Recherche de Solutions Inventives dans les Documents de Brevet par des Approches d'Apprentissage Profond

Doctorante: Xin Ni (ICUBE)
Directeur de thèse: Denis Cavallucci (ICUBE)
Encadrante: Ahmed Samet (ICUBE)
Laboratoire: ICUBE (UMR CNRS 7357)

1 Contexte de la Recherche

L'innovation est un facteur clé pour les entreprises qui développent des produits et s'engagent dans une logique de progrès continu au sein d'un secteur fortement concurrencé. Les solutions inventives, qui sont une sorte de résultats significatifs de l'innovation, peuvent être utilisées pour résoudre de manière inventive des problèmes difficiles afin de faciliter les activités de recherche et développement. En particulier, un nombre croissant de chercheurs ont remarqué que la connaissance de différents domaines industriels pourrait être utile pour élaborer des solutions inventives afin de résoudre des problèmes cibles complexes.

Aujourd'hui, dans le contexte de cette préoccupation croissante pour l'innovation en matière d'ingénierie, la demande de solutions d'ingénierie inventive a également augmenté rapidement dans les entreprises (Shirwaiker and Okudan, 2008; Jardim-Goncalves et al., 2011; Smirnov et al., 2013). En outre, l'exploration de champs de connaissances plus larges pour trouver des inspirations inventives est devenue une alternative importante pour relever les défis complexes de la fabrication (Ni, Samet, and Cavallucci, 2021). Cependant, de nombreuses entreprises s'appuient encore sur l'expérience des ingénieurs, sur des séances de remue-méninges entre différents experts ou sur la recherche classique de solutions sur Internet pour promouvoir les activités de recherche et développement. Ces méthodes obsolètes ne dorénavant plus s'adapter à la croissance actuelle du renouvellement permanent de l'information et des données dans tous les domaines.

En outre, en tant que part importante de leur stratégie, la conception inventive est devenue un facteur important pour les entreprises afin de survivre dans un contexte de concurrence exacerbée (Hao et al., 2019; Renjith, Park, and Kremer, 2020). De plus, le niveau d'innovation des produits se prête d'avantage à la fabrication ouverte (Kusiak, 2016). Bien que la plupart des ingénieurs ont compris l'importance du mélange des connaissances entre différents domaines pour la création et le développement de produits (Whiteside et al., 2009), les connaissances inventives sont toujours intrinsèquement liées aux personnes qui les

utilisent (Girodon et al., 2015) et la circulation des connaissances dans différents domaines reste difficile.

Depuis le milieu du 20e siècle, TRIZ, la théorie de la résolution de problèmes inventifs initialement proposée par Altshuller (Altshuller, Shulyak, and Rodman, 0040) après avoir analysé des millions de brevets dans tous les domaines de l'industrie, a commencé à être utilisée pour améliorer et faciliter la résolution de problèmes technologiques (Altshuller, 1999). Les modèles de TRIZ et notamment les 40 principes inventifs sont souvent utilisés pour générer des solutions inventives relatives à un problème cible. Les contradictions sont généralement contenues dans les problèmes d'ingénierie. Elles sont classées en tant que contradictions techniques ou physiques. La matrice des contradictions est un des outils les pluys répandu de TRIZ. Elle est utilisée pour résoudre les contradictions en fonction des 40 principes inventifs et des 39 paramètres du système. Comme le montre la Fig. 1, un processus classique de résolution de problèmes TRIZ est présenté.

- Le problème cible spécifique est d'abord préparé par l'utilisateur.
- À l'aide de 39 paramètres du système, TRIZ transforme le problème spécifique en un problème général.
- Trouver des solutions générales par des modèles TRIZ pour le problème général.
- Transformer les solutions générales en solutions spécifiques en appliquant les modèles TRIZ.

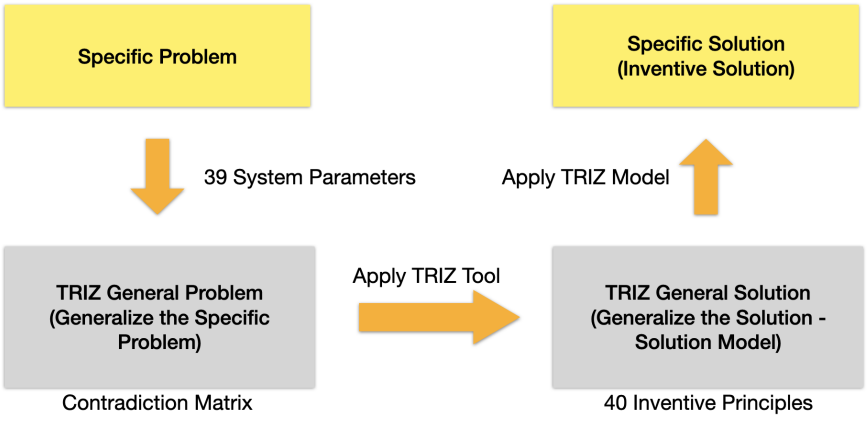

Fig. 1. Le processus de résolution de problèmes de TRIZ

Au cours des dernières décennies, un nombre croissant de travaux de recherche approfondis ont été proposés pour faciliter d'avantage la résolution des problèmes. De nouvelles techniques et de nouveles techniques sont utilisés dans ce domaine, comme les approches d'apprentissage automatique, l'exploration de brevets, les méthodes d'apprentissage profond, la conception assistée par ordinateur, etc.

En particulier, par rapport à ces techniques, les documents de brevet jouent un rôle important pour présenter les dernières connaissances inventives dans chaque domaine. En outre, les connaissances relatives à la méthode de conception inventive (IDM) qui dérive de TRIZ (Sheu, Chen, and Yu, 2012) sont principalement contenues dans les documents de brevet (Ni, Samet, and Cavallucci, 2019). Plus de 80 % des connaissances techniques de l'humanité sont décrites dans des documents de brevet (Souili, Cavallucci, and Rousselot, 2015) et l'Organisation mondiale de la propriété intellectuelle a révélé que 90% à 95 % de toutes les inventions du monde se trouvent dans des documents de brevet (Yeap, Loo, and Pang, 2003). En outre, les documents de brevet constituent des ressources intellectuelles importantes pour la protection des intérêts des individus, des organisations et des entreprises (Ni, Samet, and Cavallucci, 2019). Ils fournissent également des informations précieuses pour résoudre les problèmes d'ingénierie et renforcer l'inventivité. Les connaissances innovantes contenues dans les brevets tendent toujours à présenter les solutions les plus récentes pour résoudre les problèmes.

Par conséquent, l'utilisation efficiente et efficace des connaissances inventives existantes contenues dans les documents de brevet de différents domaines mérite d'être explorée plus avant. Néanmoins, il a toujours été difficile pour les ingénieurs qui n'ont pas une connaissance approfondie des différents domaines d'utiliser pleinement les connaissances inventives contenues dans les documents de brevet. En particulier, l'exploration de plusieurs brevets par un expert s'avère être une tâche compliquée. Il est donc important et utile d'automatiser l'ensemble du processus de recherche de solutions inventives dans les documents de brevet de différents domaines pour résoudre de manière inventive des problèmes cibles.

Aujourd'hui, divers outils et techniques d'apprentissage automatique ont été mis au point pour aider les experts à analyser les brevets et à automatiser le processus d'extraction des connaissances. En effet, les techniques de traitement du langage naturel (NLP) ont fait des progrès importants. Nous utilisons donc particulièrement les techniques NLP dans ce travail. Comme illustré dans la Fig. 2, la thèse est réalisée dans le cadre de la résolution de problèmes à partir de documents de brevets avec des techniques NLP. Elle a pour but d'aider les ingénieurs à résoudre des problèmes complexes et multidisciplinaires par le biais du processus de recherche de problèmes similaires, de mise en correspondance des problèmes et des solutions et de classement des solutions inventives. Dans ce cadre, un grand nombre de documents de brevets provenant de différents domaines sont utilisés pour extraire des solutions inventives latentes pour des problèmes réels donnés. Les connaissances inventives existantes et publiées con-

tenues dans ces documents de brevet peuvent être un outil disponible et utile pour les ingénieurs qui n'ont pas une grande compréhension des connaissances de différents domaines afin de faciliter leurs activités de R&D.

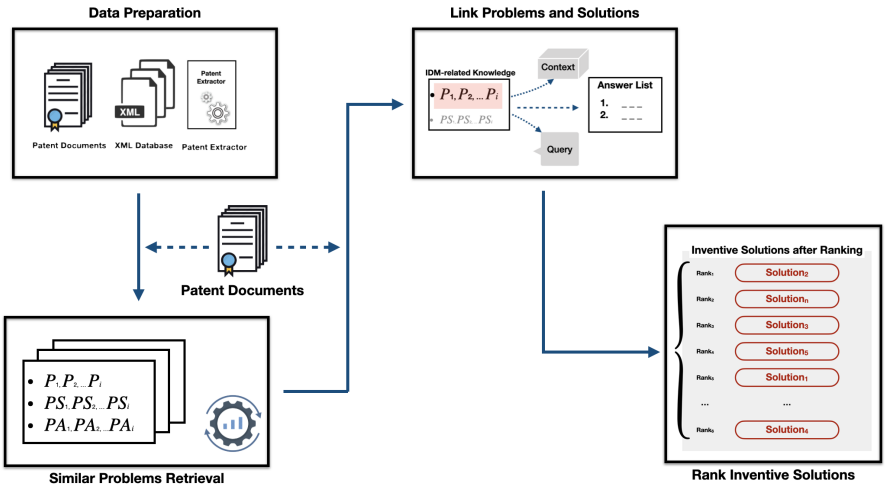

Fig. 2. L'architecture globale de la thèse

Dans cette thèse, nous nous concentrons sur le travail du processus de recherche automatique de solutions inventives complètes, qui concerne l'utilisation des connaissances inventives contenues dans différents domaines de connaissances pour résoudre de manière inventive les problèmes cibles des milieux industriels.

2 Motivation de la Recherche

La motivation de ce travail est de faciliter et d'automatiser le processus de résolution inventive de problèmes en se basant sur les technologies NLP.

Comme indiqué dans la Section 1, les documents de brevet de différents domaines contiennent la richesse des connaissances inventives publiées et les plus récentes. Ils fournissent également des informations précieuses pour résoudre les problèmes d'ingénierie et renforcer l'inventivité. Ces connaissances inventives contenues dans les documents de brevet tendent toujours à présenter les solutions les plus récentes pour résoudre les problèmes. En outre, les connaissances inventives contenues dans les brevets pourraient être définies des **problèmes** (Souili, Cavallucci, and Rousselot, 2015). Le problème décrit les caractéristiques insatisfaisantes des méthodes ou des situations existantes.

Par exemple, pour le cas d'utilisation du stylo tactile[1].

– *non-conductive material like plastic could hampers users to operate the pen with wearing gloves, having very dry skin, or some situations in which the user does not make good conductive contact with the device to the touch screen.*

Ce problème pourrait se manifester surtout lorsque l'environnement est froid. Le brevet propose donc des *solutions partielles* qui apportent des améliorations ou des changements aux problèmes définis. Une solution partielle pourrait être :

– *replacing the inner molding built by non-conductive material of touch pen with a ideally metal material device so that the stylus tip operates even when held by an extremely good insulator.*

Par conséquent, l'exploitation de ces connaissances inventives en matière de résolution de problèmes contenues dans les documents de brevet peut être utilisée pour résoudre les problèmes cibles de différents domaines lorsque ces problèmes sont suffisamment similaires.

Cependant, il existe déjà une grande quantité de documents de brevet dans le monde et de nombreux nouveaux documents de brevet sont publiés chaque année. Il est impossible de vérifier manuellement ces très nombreux documents de brevet. En outre, les documents de brevet sont répartis sur un large éventail de domaines différents. Prenons l'exemple du système de brevets classique, l'Office des brevets et des marques des États-Unis (USPTO).

Les brevets américains se répartissent en trois catégories : les brevets d'utilité, les brevets de design et les brevets de plante. Les brevets d'utilité sont accordés à toute personne qui invente ou découvre un procédé, une machine, un article manufacturé ou une composition de matière nouveaux et utiles, ou toute amélioration nouvelle et utile de ceux-ci ; les brevets de design sont accordés à toute personne qui invente un design nouveau, original et ornemental pour un article manufacturé ; les brevets de plante sont accordés à toute personne qui invente un design nouveau, original et ornemental pour un article manufacturé. Parmi eux, 90% des brevets américains sont des brevets d'utilité, qui protègent l'utilité ou les aspects fonctionnels d'une invention (Ni, Samet, and Cavallucci, 2019). En outre, le lien de l'invention des brevets d'utilité entre différents domaines est un peu plus fort que celui des brevets de design et de plantes. Ainsi, dans ce travail, nous prenons uniquement les brevets d'utilité et ils peuvent être classés en huit domaines différents : les nécessités humaines (HN), les opérations d'exécution (PO), la chimie (C), les textiles (T), les constructions fixes (FC), le génie mécanique (ME), la physique (P) et l'électricité (E). De toute évidence, pour permettre aux ingénieurs d'avoir une large compréhension de ces différents domaines, la connaissance est un obstacle crucial. Ainsi, ce travail vise à supprimer cet obstacle pour les ingénieurs afin qu'ils soient en mesure d'obtenir

[1] le lecteur peut se référer à ce lien pour le brevet complet https://patents.google.com/patent/US8847930B2/

facilement les dernières connaissances en matière de résolution de problèmes inventifs à partir de documents de brevets dans différents domaines.

En outre, nous remarquons en particulier que les documents de brevets de différents domaines peuvent contenir des problèmes sémantiques similaires. Lorsque ces problèmes sont suffisamment similaires dans la partie sémantique, leurs solutions correspondantes proposées par des documents de brevets de domaines différents pourraient être un type de solutions inventives pour le problème cible. La distance plus grande entre leurs domaines pourrait fournir des solutions latentes plus inventives ou des inspirations pour résoudre le problème cible. En particulier, la plupart des ingénieurs ne parviennent pas à maitriser les connaissances des domaines éloignés. Pour mieux utiliser ces connaissances éloignées, nous proposons un nouveau processus de recherche de connaissances inventives.

Dans le contexte de l'automatisation d'un processus complexe d'extraction de données, cela nous incite à automatiser l'ensemble du processus de recherche de solutions inventives afin d'éviter les travaux de coopération complexes entre les différentes phases de préparation. De la préparation des données à la recherche de problèmes similaires, en passant par la mise en correspondance des problèmes et des solutions et le classement des solutions inventives, ces différentes étapes coopèrent ensemble pour permettre l'automatisation de l'ensemble du processus de recherche de solutions inventives, en particulier lorsqu'un grand nombre de documents de brevets sont saisis.

Les approches d'apprentissage profond ont connu un développement rapide ces dernières années. De nombreuses approches de pointe en matière de réseaux de neurones ont obtenu des résultats prometteurs dans différents domaines et tâches de recherche, notamment dans le domaine du langage naturel. En raison de la nature de leur conception et de leur structure, les différentes approches de réseaux neuronaux présentent des performances différentes pour différentes tâches. Par exemple, pour prédire le mot cible sur la base d'informations contextuelles longues, les réseaux de mémoire à long terme (LSTM) sont plus performants que les réseaux neuronaux Word2vec. Il peut mieux apprendre les informations contextuelles plus longues autour du mot cible grâce à sa conception originale d'oubli et de portes de mémoire. En particulier, les documents de brevet sont différents des autres types de textes. Ils contiennent beaucoup de longues phrases par rapport aux autres documents. En outre, de nombreuses expressions complexes et des mots uniques sont toujours contenus dans ces longues phrases. Les informations contextuelles sont généralement plus longues et difficiles à comprendre. Tous ces éléments font que la tâche de recherche de solutions inventives sur les documents de brevets est plus unique que d'autres travaux. Cela nous incite donc à explorer différentes approches d'apprentissage profond pour répondre à notre tâche.

Dans l'ensemble, les défis susmentionnés nous incitent à proposer un processus de recherche automatique de solutions inventives. Les approches proposées visent à combiner des méthodes d'exploration de données, des technologies de

calcul de similarité sémantique et des approches d'apprentissage profond pour automatiser la recherche de solutions inventives à partir d'un grand nombre de documents de brevet. Elles peuvent fournir aux ingénieurs qui n'ont pas une connaissance approfondie de différents domaines une nouvelle façon de trouver des inspirations inventives à partir de documents de brevets pour des problèmes donnés. D'autre part, à notre connaissance, notre travail est le premier à utiliser pleinement la connaissance des documents de brevet de différents domaines pour fournir automatiquement des solutions inventives. Il facilite davantage le travail de recherche sur la résolution de problèmes dans le domaine de TRIZ.

3 Contributions de la Thèse

Dans cette thèse, nous résumons les contributions proposées dans plusieurs directions de recherche de la façon suivante:

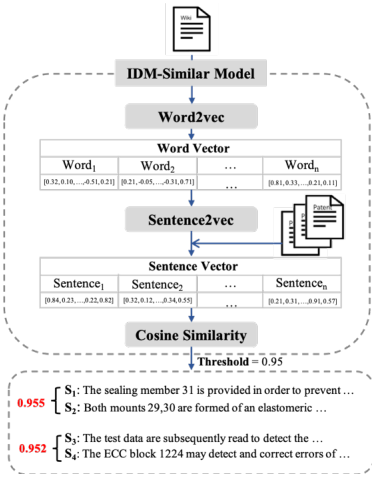

Fig. 3. Le cadre d'IDM-Similar

– **Modèle IDM-Similar basé sur les réseaux neuronaux Word2vec** : Un modèle de recherche de problèmes similaires appelé IDM-similaire est proposé par la thèse, comme le montre la Fig. 3. Selon les réseaux neuronaux Word2vec de Google (Mikolov et al., 2013), Le modèle IDM-Similar est capable d'extraire des problèmes similaires de documents de brevets de différents domaines. Il obtient le vecteur de phrase pour chaque problème cible dans les documents de brevet via Word2vec. De plus, nous formons d'abord le modèle Word2vec sur la base d'un ensemble de données Wikipedia en anglais. Il peut ainsi apprendre la similarité sémantique entre différents mots. Les représentations des phrases peuvent ensuite être réalisées

à partir des représentations des mots. La métrique de similarité Cosine est également combinée pour prédire la valeur de similarité entre les paires de phrases. Le modèle IDM-Similar est finalement capable d'extraire les phrases problématiques similaires des documents de brevet par le biais du calcul de la similarité.

– **Modèle SAM-IDM basé sur des réseaux de neurones à mémoire à long terme (LSTM)**: Afin de permettre au modèle de mieux apprendre les informations contextuelles longues en fonction des caractéristiques du document de brevet, un nouveau modèle d'extraction de problèmes similaires appelé SAM-IDM reposant sur des LSTM est proposé, comme l'illustre la Fig. 4. Il combine un modèle LSTM de Manhattan pour effectuer la tâche de comparaison de similarité sémantique entre différentes phrases. En outre, la mise en œuvre d'un processus d'élagage est utilisée pour garantir un niveau plus élevé d'inventivité et d'efficacité temporelle. Comparé au modèle IDM-Similar, SAM-IDM affiche de meilleures performances sur des documents de brevets américains des milieux industriels.

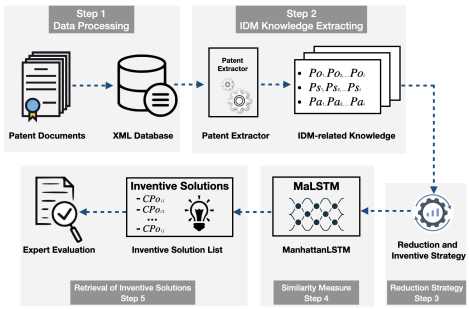

Fig. 4. Le cadre de SAM-IDM

– **IDM-Modèle de mise en correspondance basé sur les réseaux neuronaux XLNet** : Faire correspondre les problèmes et les solutions correspondantes dans les documents de brevets, un modèle appelé IDM-Matching est proposé dans la thèse, comme le montre la Fig. 5. Il combine un réseau de neurones de pointe appelé XLNet dans le domaine du langage naturel. En particulier, nous traitons cette tâche comme un système de réponse aux questions (Ravichandran and Hovy, 2002). Nous convertissons spécialement chaque problème en une requête afin d'utiliser pleinement les réseaux neuronaux XLNet et d'éviter les inconvénients des méthodes traditionnelles de correspondance lexico-syntaxique. Ce modèle vise à établir le lien entre les problèmes et les solutions dans les documents de brevet afin de faire cor-

respondre les problèmes similaires du modèle SAM-IDM avec les solutions inventives des documents de brevet de différents domaines.

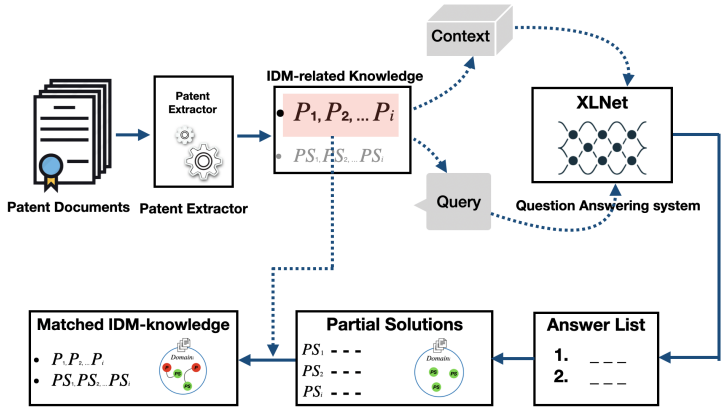

Fig. 5. Le cadre de l'IDM-Matching

- **Modèle PatRIS basé sur l'analyse de décision à critères multiples** : Un modèle appelé PatRIS basé sur l'analyse décisionnelle à critères multiples (MCDA) est proposé pour classer les solutions inventives latentes. Lorsqu'un grand nombre de documents de brevet est utilisé comme entrée du modèle SAM-IDM, plusieurs problèmes similaires peuvent être générés à partir de documents de brevet de différents domaines. Plusieurs solutions inventives latentes correspondantes peuvent donc être obtenues via le modèle IDM-Matching. Afin de mieux classer ces solutions inventives latentes, le modèle PatRIS combine une approche MCDA appelée TOP-SIS. En outre, plusieurs indicateurs d'inventivité des brevets et la valeur de similarité sont combinés avec différents poids pour construire un système de classement de l'inventivité des solutions cibles. Ce travail vise à fournir aux ingénieurs un moyen d'obtenir le plus de solutions inventives possibles lorsque le nombre de solutions inventives latentes est élevé.
- **Un démonstrateur nommé PatentSolver**: Suite aux recherches susmentionnées, un démonstrateur nommé PatentSolver est proposé dans cette thèse. Il contient les modèles décrits précédemment et vise l'automatisation de l'ensemble du processus. Plusieurs fonctions telles que la présentation des détails du brevet, la recherche du numéro de brevet, la présentation de la liste des problèmes similaires ainsi que la liste des solutions correspondantes, et le classement des solutions inventives, sont toutes développées et assemblées dans PatentSolver. Il s'agit d'un prototype du logiciel réel à fournir aux ingénieurs. Le futur logiciel peut être basé sur PatentSovler pour traiter les

données de brevet en temps réel afin de permettre au secteur industriel de bénéficier de nos travaux de recherche et de faciliter davantage les activités de R&D.

4 Publications

1. ***Ni Xin***, Ahmed Samet, Hicham Chibane, and Denis Cavallucci. "PatRIS: Patent Ranking Inventive Solutions." DEXA2021. 2021. (***Accepted***)
2. ***Ni Xin***, Ahmed Samet, and Denis Cavallucci. "Replicating TRIZ Reasoning through Deep Learning." TRIZ Future 2021. 2021. (***Accepted***)
3. ***Ni Xin***, Ahmed Samet, and Denis Cavallucci. "Similarity-based approach for inventive design solutions assistance." Journal of Intelligent Manufacturing (2021): 1-18. (***Published***)
4. Assis, Y., Nafi, A., ***Ni X.***, Samet, A., & Guarino, G. (2021). Analyse textuelle des RPQS pour la constitution de bases deconnaissances: Textual analysis of RPQS for the constitution of knowledge bases. Techniques Sciences Méthodes. 2021. (***Published***)
5. ***Ni Xin***, Ahmed Samet, and Denis Cavallucci. "Build Links Between Problems and Solutions in the Patent." International TRIZ Future Conference. Springer, Cham, 2020. (***Published***)
6. ***Ni Xin***, Ahmed Samet, and Denis Cavallucci. "Similarity Computation Supporting Creative Activities." Proceedings of the Sixth International Conference on Design Creativity (ICDC 2020). 2020. (***Published***)
7. ***Ni Xin***, Ahmed Samet, and Denis Cavallucci. "An approach merging the IDM-related knowledge." International TRIZ future conference. Springer, Cham, 2019. (***Published***)

References

Altshuller, Genrich, Lev Shulyak, and Steven Rodman (40). "principles: TRIZ keys to technical innovation. Technical Innovation Center". In: *Inc., Worcester, MA*.

Altshuller, Genrikh Saulovich (1999). *The innovation algorithm: TRIZ, systematic innovation and technical creativity*. Technical innovation center, Inc.

Girodon, Julien et al. (2015). "An organizational approach to designing an intelligent knowledge-based system: Application to the decision-making process in design projects". In: *Advanced Engineering Informatics* 29.3, pp. 696–713.

Hao, Jia et al. (2019). "An evolutionary computation based method for creative design inspiration generation". In: *Journal of Intelligent Manufacturing* 30.4, pp. 1673–1691.

Jardim-Goncalves, Ricardo et al. (2011). "Knowledge framework for intelligent manufacturing systems". In: *Journal of Intelligent Manufacturing* 22.5, pp. 725–735.

Kusiak, Andrew (2016). "Put innovation science at the heart of discovery". In: *Nature* 530.7590, pp. 255–255.

Mikolov, Tomas et al. (2013). "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781*.

Ni, Xin, Ahmed Samet, and Denis Cavallucci (2019). "An Approach Merging the IDM-Related Knowledge". In: *International TRIZ Future Conference*. Springer, pp. 147–158.

— (2021). "Similarity-based approach for inventive design solutions assistance". In: *Journal of Intelligent Manufacturing*, pp. 1–18.

Ravichandran, Deepak and Eduard Hovy (2002). "Learning surface text patterns for a question answering system". In: *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, pp. 41–47.

Renjith, Sarath C, Kijung Park, and Gül E Okudan Kremer (2020). "A design framework for additive manufacturing: Integration of additive manufacturing capabilities in the early design process". In: *International Journal of Precision Engineering and Manufacturing* 21.2, pp. 329–345.

Sheu, D Daniel, Chia-Hung Chen, and Pang-Yen Yu (2012). "Invention principles and contradiction matrix for semiconductor manufacturing industry: chemical mechanical polishing". In: *Journal of Intelligent Manufacturing* 23.5, pp. 1637–1648.

Shirwaiker, Rohan A and Gül E Okudan (2008). "Triz and axiomatic design: a review of case-studies and a proposed synergistic use". In: *Journal of Intelligent Manufacturing* 19.1, pp. 33–47.

Smirnov, Alexander et al. (2013). "Knowledge management for complex product development". In: *IFIP International Conference on Product Lifecycle Management*. Springer, pp. 110–119.

Souili, Achille, Denis Cavallucci, and François Rousselot (2015). "A lexico-syntactic pattern matching method to extract IDM-TRIZ knowledge from on-line patent databases". In: *Procedia Engineering* 131, pp. 418–425.

Whiteside, A et al. (2009). "Developing a current capability design for manufacture framework in the aerospace industry". In: *Proceedings of the 19th CIRP Design Conference–Competitive Design*. Cranfield University Press.

Yeap, Tralvex, Gim Hwa Loo, and Serena Pang (2003). "Computational patent mapping: intelligent agents for nanotechnology". In: *Proceedings International Conference on MEMS, NANO and Smart Systems*. IEEE, pp. 274–278.