INTERNATIONAL DESIGN CONFERENCE – DESIGN 2024
https://doi.org/10.1017/pds.2024.208

Towards the extraction of semantic relations in design with natural language processing

Vito Giordano [1,2,✉], Marco Consoloni [1,2], Filippo Chiarello [1,2] and Gualtiero Fantoni [1,2]

[1] University of Pisa, Italy, [2] Business Engineering for Data Science Lab (B4DS), Italy

✉ vito.giordano@unipi.it

Abstract

Natural Language Processing (NLP) has been extensively applied in design, particularly for analyzing technical documents like patents and scientific papers to identify entities such as functions, technical feature, and problems. However, there has been less focus on understanding semantic relations within literature, and a comprehensive definition of what constitutes a relation is still lacking. In this paper, we define relation in the context of design and the fundamental concepts linked to it. Subsequently, we introduce a framework for employing NLP to extract relations relevant to design.

Keywords: large language model (LLM), artificial intelligence (AI), design cognition, natural language processing (NLP)

1. Motivation

Natural Language Processing (NLP) has become an indispensable tool in design, profoundly influencing how products and services are developed. According to Siddharth et al. (2022a), NLP provides designers with deep insights by analyzing human language data, aiding in informed and customized design choices. This technology is pivotal in extracting meaningful information from various unstructured texts, including patents, academic papers, and product reviews (Chiarello et al., 2021). The field of NLP has seen significant advancements, particularly with the introduction of Large Language Models (LLMs). These models, like ChatGPT, which use deep learning to understand and process language, have dramatically changed our interaction with natural language.

NLP offers significant potential across various stages of the design process, ranging from problem definition to solution identification (Siddharth et al., 2022a). At the early stages, NLP facilitates the identification and analysis of user needs. For instance, Chiarello et al. (2020) employed sentiment analysis to interpret Amazon reviews, while Han and Moghaddam (2021) utilized the advanced NLP model BERT (Bidirectional Encoder Representations from Transformers) to extract attributes, descriptions, and sentiments from online user reviews. Furthermore, various researchers have developed NLP systems for the automatic identification of problem-solution pairs, aiding in mapping the state of the art within specific technological domains, assisting in the generation of new ideas, and studying technological evolution (Giordano et al., 2023). The integration of NLP with established design methodologies, such as design-by-analogy (Jeong and Kim, 2014) and TRIZ (Guarino et al., 2022), provides support to designers during the idea generation phase. Siddharth et al. (2022a) also highlight the role of NLP and LLMs in identifying specific design concepts, including functions (Russo and Montecchi, 2011), and technical features (Yoon and Park, 2005). Recently, there's been a growing interest in understanding and extracting semantic relations from technical documents (Han et al., 2022). This process involves identifying links between words and terms (semantic relations), providing deeper

language comprehension for design stakeholders. Utilizing semantic relations to encapsulate design knowledge offers multiple benefits, facilitating enhanced reasoning, analysis, and manipulation of the knowledge embedded within design documents. This approach significantly improves the retrieval of design information, as demonstrated by Han et al. (2022) and Sarica et al. (2023), by making the data more accessible and interpretable. Furthermore, it substantially enhances the effectiveness of all NLP applications within the design domain (Giordano et al., 2024).
The process of extracting semantic relations using NLP is referred to by various terms, such as relation extraction, semantic networks, and knowledge graphs. It's crucial to recognize that these terms, while often used interchangeably, are not identical in meaning. For instance, a knowledge graph specifically refers to a tool employed for studying semantic relations. Although there is considerable overlap in how these terms are used, each carries its distinct implications and applications.
Research in semantic relation extraction falls into two broad categories. The first focuses on identifying and categorizing relations between word pairs, while the second concentrates on merely acknowledging the existence of a relationship (Jang and Yoon, 2021). In the last class, Sarica et al. (2021; 2023) provide examples of utilizing LLMs in building semantic networks from patent data to represent design solutions, demonstrating the effectiveness of these models. However, this approach has a significant drawback: it often lacks specificity and interpretability. For what concern approaches that explicit relation types, two primary methods are employed: traditional and deep learning-based (Detroja et al., 2023). Traditional methods often use rule-based systems like Subject-Action-Object (SAO) for relation extraction. Choi et al. (2012) use SAO to classify the relations in partitive, effect and attribute type. Similarly, Guo et al. (2016) extract partitive relations to trace the technological trends. Other traditional methods involve using prepositions (of, for, by, to, from) to extract and classify relations (An et al., 2018). Jang and Yoon (2021) used a list of keywords to classify and extract relations from patents in structural (part, element, component), operational (cause, drive, input) and attributive (angle, parameter, scale) relations. Fantoni et al. (2013) develop an extensive set of list of keywords for first extract engineering design entities, such as functions, structure and components and then to relate these concepts. In contrast, deep learning approaches utilize machine learning for relation extraction. Chen et al. (2020; 2022) exemplify this approach by tagging entities and relations in patent-related sentences and developing LLM architectures for entity and relation recognition. Chen et al. (2020) built a dataset of patent-related sentences where they tagged 12 types of entities (such as system, component, function, and effect) and 11 different types of relations (such as spatial relation, part-of, and in-manner-of).
Despite these advancements, a comprehensive understanding of what constitutes a relation and its key components is crucial. Han et al. (2022) reviewed the literature on semantic networks with NLP in design, showing the different methods used to extract relations and construct semantic networks, the knowledge types taken into consideration (general, domain-specific, engineering), and defined possible research directions. However, more focus is needed on establishing theoretical frameworks to assist designers in studying relations. Without a clear theoretical frameworks, it becomes challenging to effectively extract and utilize semantic relations in design using NLP. A precise understanding of what constitutes a semantic relation is fundamental to identifying, categorizing, and applying these relations within the design context. This clarity is essential not only for the development of NLP tools and methodologies but also for enabling designers to leverage these insights in their work. Consequently, this paper aims to clarify fundamental concepts and address technical issues noted in existing literature, thereby paving the way for a deeper and more structured engagement with semantic relation extraction in the design domain.

2. Semantic relations in technical documents

2.1. Towards a definition of relation

In this section, we aim to provide a formal definition of a relation to facilitate a clear understanding of its constituent elements. Table 1 presents a set of definitions among the many found in the literature. Upon careful examination of these definitions, we can discern two essential elements necessary for defining relations. Firstly, it becomes evident that relations inherently exist in connection with pairs, sets, or multiple entities. To emphasize this concept, we have *italicized* the references to the concept of

"set" within each definition. Secondly, relations serve as the links that bind entities together. This aspect is highlighted by the use of **bold** formatting for the term "entity" in each definition. An entity represents the fundamental building block of a relation, and in the field of NLP, the process of identifying entities from textual data is referred to as Named Entity Recognition. An "entity" typically refers to a named entity, which is a term used to describe real-world objects such as persons, locations, organizations, products that can be categorized into a predefined class.
Upon closer examination of these definitions, it becomes apparent that they do not delve into the specific nature of the connections between these entities; rather, they remain somewhat abstract in their descriptions. For instance, the definition 1 characterizes a relation as "a connection", while the definitions 2, 3, and 5 emphasize that entities belong to a broader, more abstract set of entities, and the property of belonging together within this set defines the nature of a relation. The definition 4 defines a relation as "the way" entities relate. However, inserting into definition the fact that in such way the entities are linked is fundamental to extract relations using NLP. For this reason, our definition of relation is: "*A relation is the way in which a set of entities are connected*" (**Definition 1**). In summary, our exploration of these definitions reveals the fundamental elements of relations: (1) their existence in pairs or sets of entities, (2) the entities, and (3) the nature of the interaction (the way).

2.2. Types of relations in technical documents

Technical documents, such as patents and scientific publications, often contain various types of relations that contribute to the structure and understanding of the information presented. Three notable types of relations include: lexical, social, and causal relations.
Lexical relations refer to the connections between words or phrases within the text that contribute to its semantic structure. In the context of technical documents, these relations are crucial for understanding the specific terminology and concepts being discussed. For instance, in a patent document, the lexical relation might manifest in the form of synonyms used to refer to the same term or hypernymy, which identifies the relations between a term belonging to a set defined by another term. For example, in the sentence "*The energy supplier (power supplier) 207 shown in FIG. 2 supplies power to the house through the power system.*" (**CN102959339A**), the authors treat "energy supplier" and "power supplier" as synonymous terms. Different works in literature attempt to extract these relations from technical documents (An et al., 2018; Chen et al., 2022; Chen et al., 2020).

Table 1. Definitions of relation

#	Reference	Definition
1	Wikipedia (Philosophy)	A relation is a manner in which *multiple* **entities** stand to each other. It is a connection or association between **entities** and can be understood as a feature characterizing these entities as a whole.
2	Wikipedia (Mathematics)	Given a *set* X, a relation R over X is a set of ordered *pairs* of **elements** from X, formally: R ⊆ {(x,y): x,y ∈ X}.
3	Wikipedia (Database)	A relation is a *set* of *tuples* $(d_1, d_2, \cdots, d_n)$, where each **element** d_j is a member of D_j, a data domain.
4	Oxford Dictionary	A relation is the way in which *two or more* **things** are connected.
5	Jang and Yoon (2021)	A relation is defined in the form of a *tuple* $t = (e_1, e_2, \cdots, e_n)$ where e_i are **entities** in a predefined relation *r* within document *D*.
6	Siddharth et al. (2022b)	A relationship is a fact or *triple* that includes source **entity**, relationship, target **entity**.

Social relations within technical documents express the connections between different users or stakeholders of a technology. These relations provide insights into how various entities interact within the context of a given technology or system. For example, in the sentence "*Examinations, during a patient - doctor interaction, need to be intuitive, learn doctor behavior, and relegate the need to type, thereby providing a relatively faster examination recording mechanism.*" (**AU2015213496A1**), the term "patient" is related to "doctor", signifying their roles and interactions within a medical context.

Causal relations in technical documents denote the cause-and-effect relations between different components, actions, or processes (Kim and Kim, 2012; Kim et al., 2018). In patents, causal relations are frequently used to describe how various components of an invention interact and lead to a particular function or result. For instance, the sentence "*High temperature causes deterioration of the insulation value of the porcelain so that the igniter becomes short circuited.*" (**US2423809A**) establishes a causal relation between " temperature" and the degradation of "insulation value of the porcelain".

These three types of relations - lexical, social, and causal - represent only a subset of the various relations that can appear in technical documents. However, it is essential to emphasize that our primary focus is on relations that are pertinent to design knowledge and can enhance the engineering design process. Therefore, we choose not to consider lexical relations, which, while significant, are often abstract and common across all kinds of textual documents. Similarly, social relations are not taken into account, as they are only marginally connected to the field of design. In the next section, we will define and characterize these design-relevant relations in greater detail.

3. Relations in design

3.1. Atoms and molecules

As elaborated in Section 1, the extraction of relations using NLP can be categorized into two distinct approaches. The first recognizes the appearance of a semantic relation between a set of elements, while the second goes a step further by not only recognizing the existence of the relation but also attempting to classify. It is essential to emphasize that, when aiming to assign a class to a relation, knowing the classes of entities involved in the relation becomes a critical prerequisite. Without a clear understanding of the classes of entities, the task of precisely clustering relations becomes challenging, if not entirely unattainable. Furthermore, it is worth noting that even when the class of relation is recognized in the absence of well-defined entity classes, such relations often prove to be inaccurate. For instance, An et al. (2018) conducted a study in which they aimed to identify relations between "technological keywords" extracted from patents using propositions. However, the concept of a technological keyword is not a precisely defined, and it can encompass a wide range of entities including components (e.g., "battery pack" or "touchscreen device"), functions (e.g., "provide energy" or "cut material"), or technical attributes (e.g., "flow pressure" or "screen size"). These diverse categories of technological keywords can represent fundamentally different sub-aspects of technology.

An interesting observation made by the authors was related to the identification of the "inclusion" relation using the preposition "of". When technological keywords represent components, "of" aligns with the notion of inclusion. However, when dealing with technological keywords that represent functions and technical attributes, "of" does not necessarily imply an inclusion relation, but can be an axiomatic relation (Suh, 1998). For example, consider the sentences "*To improve the precision* **[technical attribute]** ***of*** *surgical drilling* **[function]**, *a technique has* …" (**US5041119A**) and "*Using the proposed device allows to reduce the diameter* **[technical attribute]** ***of*** *drilling* **[function]** *due to the use* ..." (**SU976032A1**). In these instances, the function and technical attribute are indeed connected by "of", but the relation is not one of inclusion; rather, it signifies a more complex, inherent association between the two entities.

Similarly, Choi et al. (2013) employed the Subject-Action-Object (SAO) method, focusing on partitive verbs as action (A) to extract the "inclusion" relation. In this approach, the critical factor determining the relation lies in the nature of the entities designated as the subject (S) and object (O) of the action (A). Partitive verbs, such as "include", "comprise", "have" and "supply," are chosen as indicators of inclusion relations between products or technologies. However, it's important to note that the interpretation of such verbs can be context-dependent and contingent on the specific entities involved. In the case of "supply", it can be used also as a function relation of a system (for example, when the system is a battery). These examples underscore the importance of entity categorization when interpreting relations in technical documents, highlighting the need for a context-aware approach to relation extraction.

Giordano et al. (2024) have identified several entities that frequently occur within design processes: function, behaviour, structure, physical effect, user, need, problem, solution, advantage, technical requirements, paraments and attribute. However, it is noteworthy that in their paper may not encompass

all essential design entities. Some entities, such as "misuse," "affordance," and "wirk elements" are marginally extracted with NLP, despite the existence of preliminary works attempting to recognize these entities in textual data (Chiarello et al., 2019; Melluso et al., 2021).
Another difficulty to be considered is related to certain entities can be considered atomic (a single irreducible unit) and others not. For instance, the concept "user" is an atomic entity since it represents a fundamental element in design that cannot be further decomposed into other design concepts. On the other hand, some entities are not atomic; they rely on other design concepts for their description. These non-atomic entities, by definition, involve relations between different elements. For example, Sasajima et al. (1995) and Fantoni et al. (2011) have emphasized the notion that a function is, in essence, a relation between the behaviour and the teleological information (the goal the behaviour is used for). In various studies, some of entities, such as "function," "structure," and "problem," are treated as relations, and extracted with NLP (Jang and Yoon, 2021; Fantoni et al., 2013; Chen et al., 2020). Therefore, we categorize these entities as "molecules", while the aforementioned atomic entities as "atoms".
It is essential to emphasize that both atoms and molecules can be entities within a relation or that molecules themselves can be formed by both atoms and molecules. This means that in the case of a molecule, the entity itself is a relation, and consequently, the relation connects one relation with another entity. As an example, the axiomatic relation is a relation between a function and a design parameter, and, considering a function is itself as a relation, we encounter a scenario in which a relation connects another relation with an entity.

3.2. Do design relations exist?

In the domain of design literature, numerous theories and frameworks delve into the study of relations among various design entities, explaining their significance in the design process. For instance, the Function-Behaviour-Structure (FBS) framework proposed by Gero and Kannengiesser (2004), emphasizes the roles of functions, behaviours, and structures, as well as their intricate mutual relations, in explaining how products are designed. Similarly, in TRIZ theory, problems are viewed as technical contradictions among a product's features, emphasizing the resolution of these contradictions and the manipulation of relations among them to enhance a product's ideality (Altshuller, 1984). Furthermore, the relations among technical features, such as size, weight, height, and speed, are formally structured in the TRIZ matrix, where the relations are the contradiction between pairs of technical features that can be solved with specific inventive principles. The SAPPhIRE (State Change, Action, Part, Phenomenon, Input, oRgan, Effect) framework, introduced by Chakrabarti et al. (2005), models the relations of function (via action, state change and input), behaviour (via phenomenon and effect), and structure (via organ and part). C-K theory provides a formalization of design that transcends specific domains or objects, with relations at its core (Hatchuel and Weil, 2009). These relations define the connections and interactions between the conceptual space (C) and the knowledge space (K) and represent the dynamic nature of design processes, where concepts evolve into knowledge, fostering innovation.
Practical design tools, such as Quality Function Deployment (QFD), Axiomatic Design, and Morphology Analysis, further underscore the importance of relations in design. QFD relates user needs to technical features, Axiomatic Design emphasizes relations between functions and design parameters, and Morphology Analysis links attributes and functions.
Several works in the literature focus on extracting and identifying these design relations using NLP. Fantoni et al. (2013) construct a knowledge graph of the FBS model using patent documents, while Vincente-Gomila et al. (2021) attempt to identify TRIZ contradictions in scientific papers, exploring the evolution of dye-sensitized solar cell technology. Yoon and Park (2005) employ Morphology Analysis and NLP to identify functions, attributes, and their relations from patents, seeking novel technological opportunities. In this domain, where relations play a pivotal role and numerous entities are interconnected, a fundamental question emerges: "Can we classify these relations into a few distinct classes?". Some researchers have attempted to classify semantic relations appearing in technical documents. For instance, Jang and Yoon (2021) identify three macro classes: structural, operational, and attribute viewpoints. Chen et al. (2020) delve deeper, identifying 13 types of semantic relations, such as spatial, causative, and teleological, which provide a more detailed classification. However, it's important to note that some relations may be context-dependent and are not always captured by these

classifications. For instance, none of the relations identified by Jang and Yoon (2021) encompass the relations explained in TRIZ theory, in QFD or in Axiomatic Design.

4. NLP for relations extraction

4.1. Explicit and implicit

Semantic relations in text can be categorized into two main types: explicit and implicit relations. In accordance with **Definition 1**, explicit relations are those where both the entities involved and the way in which they are linked are clearly articulated within the sentence. Conversely, implicit relations are those in which either the entities, the way of relation, or both are not explicitly stated in the sentence.
In the sentence, "*Implementation of the vapor pressure method has been difficult due to its reliance on a relatively small pressure difference, resulting in unstable power output.*" (**US2012135282A1**), the relation between the design parameter (pressure) and the attribute (power output) as well the nature of their relation (resulting) are clearly expressed in the sentence. In the sentence, "*However, the use of these knives forces the operator to perform considerable efforts, especially in view of the fact that the meat to be processed is stored in refrigerating rooms at low temperature and therefore offers considerable resistance to cutting.*" (**US2004088866A1**), the relation between the technical attribute (temperature of the room) and the function (cutting) is not directly stated in the sentence. Instead, the relationship is understood through context and inference. Here, the sentence implies that the low temperature of the refrigerating rooms makes the meet freeze and poses a challenge to the cutting function, but it does not explicitly state this relation.
Implicit relations are a manifestation of the cognitive processes that humans employ to read a document and more in general in comprehending the world. By leveraging their cognitive resources, humans can go beyond the explicit information presented in a text or context, filling in gaps and making educated inferences about the relations between entities or concepts. Human cognition frequently involve drawing upon a vast set of background information, common sense, and domain-specific knowledge to make inferences and derive meaning. Indeed, the existence of implicit relations in a sentence or text is intricately linked to the human knowledge, which is mainly expressed in written form. This suggests that even if implicit relations are not explicitly detailed within the same document, they may clearly be expressed in explicit format in other documents.
Previous research efforts have mainly focused on the extraction of explicit relations using NLP (Detroja et al., 2023). However, in practical scenarios, entities within a text often exhibit their interrelations across multiple sentences, leading to instances of implicit relations. Although some studies have put forward methodologies for cross-sentence relation extraction (Sun et al., 2022a) to address these implicit relations, there is a need for further research to enhance the effectiveness of these existing approaches.
The challenge becomes notably more complex when attempting to capture implicit relations that are not explicitly expressed within a single document but are dispersed throughout the extensive human knowledge that exists in written form. LLMs, such as word2vec, BERT, or GPT3, have the capability to recognize the presence of relations between two entities within a document also when the relation is implicit. However, LLMs still face significant limitations in understanding the specific type and class of these relations. This challenge remains an ongoing issue.

4.2. Roadmap and timeline towards relations extraction

As we delve into the intricate task of extracting relations from technical documents using NLP, it is paramount to adopt a systematic approach. To facilitate this, we present a structured roadmap in Figure 1, outlining the sequential steps necessary to guide future research in this domain. Furthermore, in Table 2, we present a set of research questions corresponding to each phase of this roadmap. For each research question we suggest a method that can be adopted to address the topic. The method specifies also the approach, distinguishing between theoretical and empirical, and the sources. Table 2 serves as a starting point of a broader range of research questions to address for enriching the field of relation extraction using NLP, and it does not limit other possible approaches and research questions. The roadmap detailed in Figure 1 starts with the step of (1) defining the core entities, we called as "atoms" and "molecules" in Section 3 that are central to the domain of design.

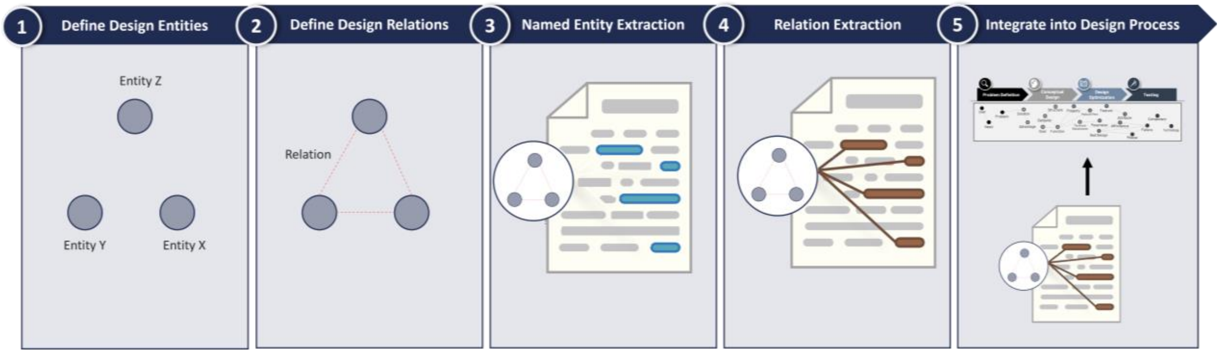

Figure 1. Roadmap towards Relations Extraction in Design

How we stress in all papers, these entities serve as the building blocks for subsequent relation extraction efforts. Once the entities are established, the next phase involves (2) categorizing and classifying the various types of relations that exist among these entities within the context of design. We believe that addressing the research questions related to (1) and (2) necessitates a two-pronged approach. Firstly, we suggest a theoretical approach that involves the development of a comprehensive framework, drawing upon existing literature (referencing research questions A, B, C). Secondly, we propose the use of empirical methods (related to D), which entails a difference-in-differences analysis on patent data. This analysis aims to discern any changes in semantic relations which we can observe from textual data of patents following the advent of artificial intelligence.

In phase (3), the research shifts its focus towards extracting entities from textual data through NLP and in particular using NER techniques. Following entity extraction, the attention turns to (4) the extraction of relations from the technical documents. This step involves leveraging NLP to unveil relations between the identified entities, with an emphasis on capturing not only explicit but also implicit relations. In these phases (3, and 4) empirical method is employed to evaluate the performance of NLP systems in comparison to existing ones (G) and in comparison with the designer requirements (F). Additionally, we emphasize the importance of developing a novel architecture to investigate how designer knowledge can be integrated to extract atoms, molecules, and relations (E). To tackle this issue effectively, a systematic literature review of the current architectures of relation extraction and their performance is essential.

Lastly (in phase 5), we underscore the need to apply and evaluate the NLP systems developed in practical scenarios and to establish novel metrics for measuring their impact. Existing literature inadequately addresses this challenge, and we contend that different metrics may be applicable for each phase of the design. Consequently, we propose a comprehensive examination of the state-of-the-art measurements utilized in design to assess specific design tasks. This analysis will serve as a foundational step towards defining the measurement for evaluating NLP in the context of each specific task (L).

For each research question in Table 2, defining a case study alongside the proposed method is crucial, with clear benchmarks for performance measurement and comparison. For instance, for research question A, a text analysis using NLP could identify key entities in famous design texts. Such an analysis might reveal, for example, that Pahl et al. (2007) clearly defines structures and functions, while Suh (1998) thoroughly explains functional requirements and design parameters along with axiomatic relationships. Similarly, TRIZ theory is expected to detail problems, solutions, and design parameters. By amalgamating these findings and analysing their occurrences and distributions, a theoretical framework can be established to encapsulate the entities of design.

5. Conclusions

There are several practical applications for extracting relation in design using automatic tools like NLP. Firstly, these systems can significantly aid designers during the idea generation phase, serving as support tools to provide design stimuli (Sarica et al., 2021; Jiang et al., 2023), or aiding in design abstraction to expand the design space and foster innovative solutions (Sun et al., 2022b). Additionally, relation extraction is crucial for visually representing design descriptions through semantic networks. This enhances various aspects of the design process, including learning, analysis, communication, conceptualization, cognition, and the redesign of design artifacts (Sarica et al., 2023). The study of relations can also improve prior art search systems in patent databases, offering designers a solid knowledge base for better problem framing and solution definition (An et al., 2021). For R&D and innovation managers, the systems can be instrumental in tracking technological trends (Yoon and Kim, 2012) and identifying new technology opportunities (An et al., 2018), as well as in conducting technology planning (Choi et al., 2012).

The innovative points of this paper are: (1) It clarifies the definition of key semantic relation concepts distinguishing between three main concepts - existence in sets of entities, the entities, and the nature of the interaction; (2) it prioritizes relations pertinent to design knowledge, focusing on those that directly enhance the engineering design process over more general or abstract relations; (3) it underscores the importance of the entities involved in these relations and the nature of their interactions, which is critical for effectively extracting and utilizing design knowledge. Furthermore, the paper delineates a roadmap to guide and advance future studies in this topic, laying the groundwork for the integration of NLP in design processes.

Table 2. Research questions for each phase of the roadmap

# Phase	**Code**	**Research Questions**	**Method (Source) - *Approach***
1	A	What are the fundamental atoms and molecules that constitute the core of engineering design, and how do these entities interact to form complex engineering systems?	Framework Development (**Design Books**) - *Theoretical*
	B	How do specific elements (atoms and molecules) influence the engineering design process, and what role do they play in determining the functionality of engineering products?	Systematic Literature Review (**Scientific Publications on Design**) - *Theoretical*
2	C	What are the different types of relations that exist between the atoms and molecules of engineering design, and how can they be systematically classified?	Systematic Literature Review (**Scientific Publications on Relation Extraction and Design**) - *Theoretical*
	D	How have recent technological advancements (like AI, machine learning) altered or enhanced the relationships between these entities in engineering design?	Difference-in-Differences analysis (**Patents**) - *Empirical*
3	E	How can design relation knowledge support NER systems in recognizing atoms and molecules?	Novel NER Architecture (**Scientific Publications on NLP**) - **architecture** *Theoretical*
	F	How we can use novel generative LLMs model, like GPT3, to facilitate the extraction of design entities?	Validation with designers - *Empirical*
4	G	How we can use NLP to capture and classify implicit relations?	Benchmark with previous methods (**Scientific Publications on Relation Extraction**) - *Empirical*
	H	Can LLMs also uncover new relations unknown to design researchers, thereby contributing to advancements in the theoretical literature of this field?	NLP analysis on outliers (**Patents, Product Descriptions**) - *Empirical*
5	I	Which stages of the design process benefit most from the integration of LLM-based relation extraction, and why?	Interview with designers - *Empirical Approach*
	L	What are the most appropriate metrics to assess the efficiency, and overall effectiveness of relation extraction models in the design phase?	Systematic Literature Review (**Scientific Publications on Design**) - *Theoretical*

The application of NLP in the field of design holds immense potential. However, there is significant work to be done to effectively utilize these systems in practical applications. A crucial question arises: can we navigate through the "black box" nature of these models to bring clarity to the complex field of design where domain knowledge is fundamental? This challenge underscores the need for continued research and development in this area to fully harness the capabilities of NLP in practical design applications.

Acknowledgement

This research has been partly funded by PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", funded by the European Commission under the NextGeneration EU programme.

References

Altshuller, G. S. (1984). "Creativity as an exact science: the theory of the solution of inventive problems".

An, J., Kim, K., Mortara, L., & Lee, S. (2018). "Deriving technology intelligence from patents: Preposition-based semantic analysis". *Journal of Informetrics*, 12(1), 217-236. https://doi.org/10.1016/j.joi.2018.01.001

An, X., Li, J., Xu, S., Chen, L., & Sun, W. (2021). "An improved patent similarity measurement based on entities and semantic relations". *Journal of Informetrics*, 15(2), 101135. https://doi.org/10.1016/j.joi.2021.101135

Chakrabarti, A., Sarkar, P., Leelavathamma, B., & Nataraju, B. S. (2005). "A functional representation for aiding biomimetic and artificial inspiration of new ideas". *AI EDAM*, 19(2), 113-132. https://doi.org/10.1017/S0890060405050109

Chen, L., Xu, S., Zhu, L., Zhang, J., Lei, X., & Yang, G. (2020). "A deep learning based method for extracting semantic information from patent documents". *Scientometrics*, 125, 289-312. https://doi.org/10.1007/s11192-020-03634-y

Chen, L., Xu, S., Zhu, L., Zhang, J., Yang, G., & Xu, H. (2022). "A deep learning based method benefiting from characteristics of patents for semantic relation classification". *Journal of Informetrics*, 16(3), 101312. https://doi.org/10.1016/j.joi.2022.101312

Chiarello, F., Cirri, I., Melluso, N., Fantoni, G., Bonaccorsi, A., & Pavanello, T. (2019). "Approaches to automatically extract affordances from patents". In *Proceedings of the Design Society: International Conference on Engineering Design* (Vol. 1, No. 1, pp. 2487-2496). Cambridge University Press. https://doi.org/10.1017/dsi.2019.255

Chiarello, F., Bonaccorsi, A., & Fantoni, G. (2020). "Technical sentiment analysis. Measuring advantages and drawbacks of new products using social media". *Computers in Industry*, 123, 103299. https://doi.org/10.1016/j.compind.2020.103299

Chiarello, F., Belingheri, P., & Fantoni, G. (2021). "Data science for engineering design: State of the art and future directions". *Computers in Industry*, 129, 103447. https://doi.org/10.1016/j.compind.2021.103447

Choi, S., Park, H., Kang, D., Lee, J. Y., & Kim, K. (2012). "An SAO-based text mining approach to building a technology tree for technology planning". *Expert Systems with Applications*, 39(13), 11443-11455. https://doi.org/10.1016/j.eswa.2012.04.014

Choi, S., Kim, H., Yoon, J., Kim, K., & Lee, J. Y. (2013). "An SAO-based text-mining approach for technology roadmapping using patent information". *R&D Management*, 43(1), 52-74.

Detroja, K., Bhensdadia, C. K., & Bhatt, B. S. (2023). "A Survey on Relation Extraction". *Intelligent Systems with Applications*, 200244. https://doi.org/10.1016/j.iswa.2023.200244

Fantoni, G., Apreda, R., Gabelloni, D., & Bonaccorsi, A. (2011). "Do functions exist?". In *DS 68-2: Proceedings of the 18th International Conference on Engineering Design (ICED 11)*, Impacting Society through Engineering Design, Vol. 2: Design Theory and Research Methodology, Lyngby/Copenhagen, Denmark, 15.-19.08. 2011 (pp. 304-313).

Fantoni, G., Apreda, R., Dell'Orletta, F., & Monge, M. (2013). "Automatic extraction of function–behaviour–state information from patents". *Advanced Engineering Informatics*, 27(3), 317-334. https://doi.org/10.1016/j.aei.2013.04.004

Gero, J. S., & Kannengiesser, U. (2004). "The situated function–behaviour–structure framework". *Design studies*, 25(4), 373-391. https://doi.org/10.1016/j.destud.2003.10.010

Giordano, V., Puccetti, G., Chiarello, F., Pavanello, T., & Fantoni, G. (2023). "Unveiling the inventive process from patents by extracting problems, solutions and advantages with natural language processing". *Expert Systems with Applications*, 229, 120499. https://doi.org/10.1016/j.eswa.2023.120499

Giordano, V., Chiarello, F., Fantoni, G. (2024)."How Engineering Design can Enhance Tech Mining: a Bibliometric and Systematic Literature Review". *Technological Forecasting and Social Change* [Second Round Review]

Guarino, G., Samet, A., & Cavallucci, D. (2022). "PaTRIZ: A framework for mining TRIZ contradictions in patents". *Expert Systems with Applications*, 207, 117942. https://doi.org/10.1016/j.eswa.2022.117942

Guo, J., Wang, X., Li, Q., & Zhu, D. (2016). "Subject–action–object-based morphology analysis for determining the direction of technological change". *Technological Forecasting and Social Change*, 105, 27-40. https://doi.org/10.1016/j.techfore.2016.01.028

Hatchuel, A., & Weil, B. (2009). "CK design theory: an advanced formulation". *Research in engineering design*, 19, 181-192. https://doi.org/10.1007/s00163-008-0043-4

Han, Y., & Moghaddam, M. (2021). "Eliciting attribute-level user needs from online reviews with deep language models and information extraction". *Journal of Mechanical Design*, 143(6), 061403. https://doi.org/10.1115/1.4048819

Han, J., Sarica, S., Shi, F., & Luo, J. (2022). "Semantic networks for engineering design: state of the art and future directions". *Journal of Mechanical Design*, 144(2), 020802. https://doi.org/10.1115/1.4052148

Jang, H. and Yoon, B., (2021). "TechWordNet: Development of semantic relation for technology information analysis using F-term and natural language processing". *Information Processing & Management*, 58(6), p.102752. https://doi.org/10.1016/j.ipm.2021.102752

Jeong, C., & Kim, K. (2014). "Creating patents on the new technology using analogy-based patent mining". *Expert Systems with Applications*, 41(8), 3605-3614. https://doi.org/10.1016/j.eswa.2013.11.045

Jiang, P., Atherton, M., & Sorce, S. (2023). "Extraction and linking of motivation, specification and structure of inventions for early design use". *Journal of Engineering Design*, 1-26. https://doi.org/10.1080/09544828.2023.2227934

Kim, H., & Kim, K. (2012). "Causality-based function network for identifying technological analogy". *Expert Systems with Applications*, 39(12), 10607-10619. https://doi.org/10.1016/j.eswa.2012.02.156

Kim, H., Joung, J., & Kim, K. (2018). "Semi-automatic extraction of technological causality from patents". *Computers & Industrial Engineering*, 115, 532-542. https://doi.org/10.1016/j.cie.2017.12.004

Melluso, N., Pardelli, S., Fantoni, G., Chiarello, F., & Bonaccorsi, A. (2021). "Detecting bad design and bias from patents". *Proceedings of the Design Society*, 1, 1173-1182. https://doi.org/10.1017/pds.2021.117

Pahl, G., Beitz, W., Feldhusen, J., & Grote, K. H. (2007). Engineering Design: A Systematic Approach. Engineering Design: A Systematic Approach.

Russo, D., & Montecchi, T. (2011). "A function-behaviour oriented search for patent digging". In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 54792, pp. 1111-1120).

Sarica, S., Song, B., Luo, J., & Wood, K. L. (2021). "Idea generation with technology semantic network". *AI EDAM*, 35(3), 265-283. https://doi.org/10.1017/S0890060421000020

Sarica, S., Han, J., & Luo, J. (2023). "Design representation as semantic networks". *Computers in Industry*, 144, 103791. https://doi.org/10.1016/j.compind.2022.103791

Sasajima, M., Kitamura, Y., Ikeda, M., & Mizoguchi, R. (1995). "FBRL: A function and behavior representation language". In *IJCAI* (Vol. 95, pp. 1830-1836).

Siddharth, L., Blessing, L., & Luo, J. (2022a). "Natural language processing in-and-for design research". *Design Science*, 8, e21. https://doi.org/10.1017/dsj.2022.16

Siddharth, L., Blessing, L. T., Wood, K. L., & Luo, J. (2022b). "Engineering knowledge graph from patent database". *Journal of Computing and Information Science in Engineering*, 22(2), 021008. https://doi.org/10.3390/pr11102831

Suh, N. P. (1998). "Axiomatic design theory for systems". *Research in engineering design*, 10, 189-209. https://doi.org/10.1007/s001639870001

Sun, Q., Xu, T., Zhang, K., Huang, K., Lv, L., Li, X., ... & Dore-Natteh, D. (2022a). "Dual-Channel and Hierarchical Graph Convolutional Networks for document-level relation extraction". *Expert Systems with Applications*, 205, 117678. https://doi.org/10.1016/j.eswa.2022.117678

Sun, Y., Liu, W., Cao, G., Peng, Q., Gu, J., & Fu, J. (2022b). "Effective design knowledge abstraction from Chinese patents based on a meta-model of the patent design knowledge graph". *Computers in Industry*, 142, 103749. https://doi.org/10.1016/j.compind.2022.103749

Yoon, B., & Park, Y. (2005). "A systematic approach for identifying technology opportunities: Keyword-based morphology analysis". *Technological Forecasting and Social Change*, 72(2), 145-160. https://doi.org/10.1016/j.techfore.2004.08.011

Yoon, J., & Kim, K. (2012). "An analysis of property–function based patent networks for strategic R&D planning in fast-moving industries: The case of silicon-based thin film solar cells". *Expert Systems with Applications*, 39(9), 7709-7717. https://doi.org/10.1016/j.eswa.2012.01.035