



HAL
open science

Using NLP to Detect Tradeoffs in Employee Reviews

Justus Schollmeyer

► **To cite this version:**

Justus Schollmeyer. Using NLP to Detect Tradeoffs in Employee Reviews. 22th International TRIZ Future Conference (TFC), Sep 2022, Warsaw, Poland. pp.209-219, 10.1007/978-3-031-17288-5_19. hal-04449786

HAL Id: hal-04449786

<https://inria.hal.science/hal-04449786v1>

Submitted on 15 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Using NLP to Detect Tradeoffs in Employee Reviews

Justus Schollmeyer¹

¹ University of Bremen, 28359 Bremen, Germany
justusschollmeyer@gmail.com

Abstract. This paper presents a methodology to identify industry-specific tradeoffs using natural language processing (NLP) to analyze employee reviews from Glassdoor. The analysis is based on 400,000+ reviews from employees working in the financial sector between 2008 and 2020. For each review, the pros and cons sections are classified in a one-to-many approach in terms of the most prevalent topics in the sector. The most prevalent noun chunks within a representative sample of reviews are used as topics. The classification of reviews is based on the cosine similarity between the sentence embeddings of these topics and the sentence embeddings of the comment sections using a sentence-transformer model. Based on this classification, the count of pro-con topic pairs is tested for statistical significance against a control of randomly generated pairings from the same sample of classifications. The process is repeated 10,000 times and only pairs with a p-value $< .05$ (after Bonferroni-Holm correction) are considered. If both, a pro-con pair of topics and its opposite pairing are significantly frequent, this combination of topics qualifies as a tradeoff. Depending on whether the same pairing constitutes a significantly frequent pro-pro pair, it can be determined whether a solution for the tradeoff exists in the sector (at least within the limitations of this approach). Using 13 topic labels, 2 tradeoffs – *work-life balance vs. opportunity* (without solution in the sector) and *work-life balance vs. management* (with solution in the sector) – have been identified. In addition to that, 8 more pro-con pairs were identified as significant.

Keywords: Tradeoffs, Glassdoor, Employee reviews, Natural Language Processing, Machine Learning.

1 Introduction

The goal of this paper is to demonstrate that the free text in employee reviews from online platforms such as Glassdoor [1] or Indeed [2] can be used to identify sector-specific tradeoffs that organizations face from the perspective of their employees. For this purpose, 439,599 Glassdoor reviews written between 2008 and 2020 by employees from the financial industry in the US will be processed using NLP.

The paper starts with a discussion of why tradeoffs matter to innovation (2.1) followed by brief introductions to employee reviews on Glassdoor (2.2) and the use of NLP for their analysis (2.3). Section 3 will introduce the method of using the pros and cons sections to identify significantly frequent pro-con pairs and their opposites.

Section 4 presents the results followed by a brief discussion of the findings and how to improve the approach further (section 5).

2 Background

2.1 Why tradeoffs matter to innovation

Tradeoffs are problems with conflicting targets. Improving one results in the deterioration of the other and vice versa. Problems that present themselves in this form bear high potential for innovation [3]. This is because to improve the values for both conflicting parameters, it is required to rethink the structure of the system (or some aspects of it).

A simple example: A container to deposit trash is 300m away from someone's house. The trash gets collected in a bin in the kitchen. The larger the bin, the fewer meters must be walked to deposit the trash over a given period. Increasing the bin size comes at the cost of a heavier bin once filled with trash. "Meters walked to the container in a month" can be seen as a negatively sloped function of the filled bin's mass if everything else remains stable (see left side of Fig.1). The function defines the options if only the size of the bin can be changed.

If, however, someone decided to go beyond these artificial limitations, they might come up with a solution that improves both parameters to the better, which can come in two forms: The system is modified such that

- (i) the entire function shifts to the better for both parameters, or
- (ii) the tradeoff disappears entirely.

Type (i) is a more general version of Schumpeter's definition of innovation: a shift of the production function [4]. Type (ii) is even preferable (at least as far as these parameters are concerned) and lies on the vector of the ideal solution defined in [3] as *function without a system*.

For both types of solutions, creative thinking is required since mental limitations – such as only allowing ourselves to change the size of the bin – must be abandoned. One could e.g., decide to produce less trash. This would shift the function to the better in the sense of (i) (see right side of Fig.1). One could also decide to integrate the walk to the container into one's daily routines, such as going to work, grocery shopping, jogging, etc. The problem becomes obsolete (ii) since no extra time is being spent on walking to the container if it is on the way of other errands. Unsurprisingly, most people apply this strategy to deal with the tradeoff.

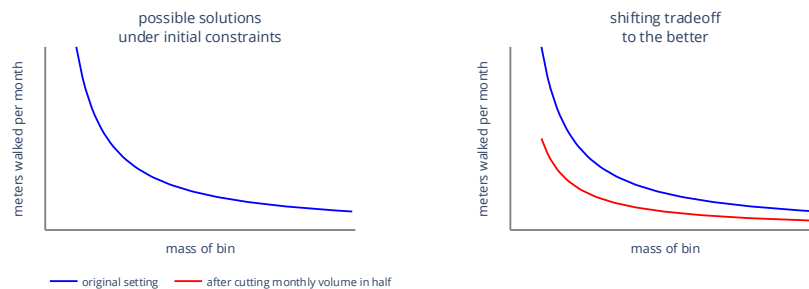


Fig. 1. The left side figure shows the meters walked per month as a function of the mass of the bin. The graph assumes that other variables such as type of trash, volume of trash, distance to the container, etc. remain stable. The graph to the right shows how the curve changes if the volume of trash is cut in half.

We cannot know for certain that a tradeoff is not resolvable in principle (because we cannot be certain that all our assumptions are true and exhaustive). However, we can check whether a solution to what looks like a tradeoff already exists. If none exists and we find a type (i) or (ii) solution, we have made a discovery or invention that could result in innovation if implemented at scale [4].

The approach of analyzing employee reviews from the financial sector presented in this paper will show how to distinguish tradeoffs that have a solution in the sector from those that seem to have no solution yet.

2.2 Employee reviews on Glassdoor

Glassdoor [1] is a website that allows current and former employees to review their companies by leaving ratings for *career opportunities*, *compensation & benefits*, *work-life balance*, *senior management*, and *culture & values*. In addition to that, employees can decide whether they approve of the CEO, whether they would recommend the company to a friend, and how they feel about the business outlook for the next six months. Finally, they are asked to write about the pros and the cons and to provide feedback to management. The analysis in this paper makes use of the pros and cons structure in the free text section. For a detailed description of the method see the methods section (section 3) and the code on GitHub [5].

2.3 NLP for analyzing Glassdoor reviews

Recent advancements in natural language processing (NLP) have proven useful in *human resource management* allowing for the analysis of qualitative data without the need for laboriously coding text into categories [6]. Both, bag-of-words approaches (e.g., [7]) and approaches based on vector-representations (e.g., [8]) have been used successfully on Glassdoor data. This article relies on both of them: a bag-of-words approach combined with part of speech tagging to identify the most prevalent noun chunks that will serve as categories, and sentence-embeddings to compute the cosine-

similarity between the categories and the comments using a sentence-transformer model [9], which maps paragraphs and sentences to a 768 dimensional dense vector space [10].¹

3 Methods

For more details see the implementation of the approach on GitHub [5].

3.1 The data

The analysis is based on 439,599 comments from the financial sector featuring 12,868 companies² submitted between April 23rd, 2008, and September 8th, 2020, to Glassdoor.

3.2 The labels

To identify the most prevalent categories, 4,000 comments have been sampled from the *pros* and the *cons* sections, respectively. The resulting 8,000 comments were used to identify the 15 most prevalent noun chunks using spaCy [11] (ignoring stop words and removing adjectives by means of spaCy’s parts of speech tags). Their number has been further reduced by removing semantic duplicates. Duplicates were determined in terms of the cosine similarity between the categories’ semantic vector representations. The vector representations were obtained by means of the *paraphrase-distilroberta-base-v1* transformer model from the Huggingface library [10]. Categories with cosine similarities above .7 were considered as duplicates. To cluster the similarity scores, the similarity matrix is transformed to a distance matrix (i.e., a matrix of dissimilarity scores). Similar topics are identified by means of hierarchical clustering and a dissimilarity threshold of .3, which corresponds to a similarity score of .7. Among such semantic duplicates, the more frequent category was kept unless it contained lemmata appearing in other topics. In the case of the latter, the less frequent category was kept. For example, the topics “work” and “job” are duplicates according to this definition. “Work” is more frequent than “job”. However, “work” also appears in the topic “work life balance”. “Job” is thus kept and “work” discarded. This procedure reduced the 15 categories down to 13.

¹ This model is also used to reduce the number of categories by removing semantic duplicates (see section 3).

² The companies comprise organizations from the following industries: Financial Transaction Processing, Investment Banking & Asset Management, Banks & Credit Unions, Lending, Brokerage Services, Financial Analytics & Research, Stock Exchanges, Venture Capital & Private Equity.

3.3 Mapping comments to labels

To classify the comments, the same model used for label reduction (see 2.2) is used to generate vector representations [10]: The vector representation of each comment is compared to the vector representation of each label in terms of their cosine similarity. If the similarity score is larger than .35, the label is assigned to the comment.

Three more labels are added to the existing list: *nothing negative to say*, *nothing positive to say*, *none*. Their only role is to identify comments that are to be removed because they do not carry relevant information for the purpose of identifying tradeoffs.

3.4 Identifying tradeoffs

The general idea behind the methodology to identify tradeoffs is described in Fig. 2. Each review has a *pros* and a *cons* section. These sections are assigned to category names (i.e., labels) in a one-to-many approach (see 2.2 and 2.3). The procedure results in a list of pro-con pairs (the maximal count of pairs is equal to the number of labels squared; in this case, 169 pairs). Pairs with the same label in the pros and the cons section are discarded (leaving maximally 156 pairs).

Statistical significance of the count of pairs is tested against a control obtained by randomly reshuffling the pairings 10,000 times. A pro-con pair counts as significant, if the Holm-Bonferroni adjusted p-value [12] is below $< .05$ when compared to the randomly created distribution.³ If both, a pro-con pair and its opposite meet the significance test, the combination qualifies as a tradeoff (see Fig. 2).

³ The more pairs tested for statistical significance, the higher the probability of obtaining type I errors (false positives). To control for this without unnecessarily increasing the risk of type II errors (false negatives), the Holm-Bonferroni method is used to adjust the family-wise error rates (FWER) for each pair. The p-values are sorted and ranked from lowest to highest starting with rank 0. For each pair that is to be tested, the FWER (in this case .05) is then divided by the number of tests (in this case 156) minus the pair's rank. This results in an array of adjusted FWERs. If an adjusted FWER is smaller than its predecessor, it is set equal to the rate of the previous pair. A pair is significant if its p-value is lower than its adjusted FWER.

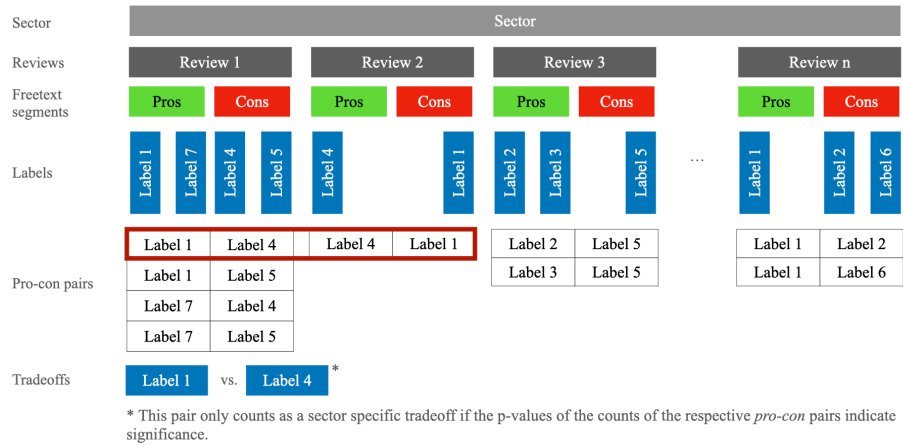


Fig. 2. The pros and cons sections of employee reviews are classified in terms of categories (labels) relevant to the sector. Only pairs of labels that are significantly frequent qualify as pro-con pairs. If the opposite of the pro-con pair constitutes another significantly frequent pro-con pair, the combination counts as a tradeoff.

To evaluate whether a sector has a solution for a tradeoff, the same procedure is carried out for the labels within the *pros* and the *cons* sections, respectively. The goal is to check whether the *either-or* combination also exists in the form of *both-both*. For example, if significantly many reviewers report *good work-life balance* and *bad management*, and another group reports the opposite, it could still be the case that another group of reviewers reports both *good work-life balance* and *good management*. The latter is called a *pro-pro* pair. Tradeoffs whose labels also appear significantly often as pro-pro pairs count as tradeoffs that have a solution within the sector. In contrast, if a pro-pro combination of the labels constituting a tradeoff is significantly infrequent, it can be assumed to have no obvious solution within the sector – at least not within the constraints of this approach (see Fig. 3).

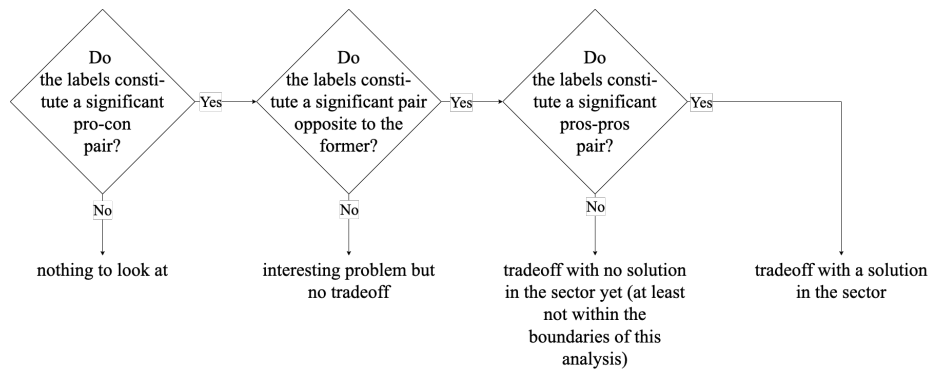


Fig. 3. Illustration of the decision logic explaining the difference between tradeoffs with a solution in the sector and tradeoffs without such a solution (in the boundaries of this analysis).

4 Results

4.1 Tradeoffs

Filtering for the most prevalent nouns (noun chunks included) results in the following list of 13 categories used as labels: *company*, *people*, *management*, *employee*, *benefit*, *time*, *job*, *pay*, *opportunity*, *culture*, *hour*, *training*, and *work life balance*.

Running the significance test on the total counts of all pairings reveals two tradeoffs (see Table 1): *management vs. work life balance* (870, $P < .0001$ and 118, $P < .0001$), and *opportunity vs. work life balance* (47, $P < .0001$ and 54, $P < .0001$). The labels *management & work-life balance* appear significantly frequently together in the pros section as a *pro-pro* pair, which suggests that a solution exists within the sector (145, $P < .0001$). In contrast, the pair *opportunity & work life balance* is significantly infrequent (11, $P < .0001$) and thus has no obvious solution in the sector (at least not within the boundaries of this analysis).

Table 1. Summary of the tradeoffs.

con	pro	count	p-val (adj.)	cons pairing	pros pairing
management	work life balance	870	0		
work life balance	management	118	0	insignificant	frequent
opportunity	work life balance	47	0		
work life balance	opportunity	54	0	infrequent	infrequent

For an illustration of the tradeoff ratios within firms, see Fig. 4.



Fig. 4. Illustration of the ratios between the two sides of the respective tradeoffs within companies. The y-axis (logarithmic scale) shows the number of companies that fall into the respective ratio bins.

For examples of reviews illustrating the tradeoffs, see table 2.

Table 2. Examples of reviews illustrating the tradeoffs.

label cons	label pros	Cons	pros
work life balance	opportunity	Not the best work life balance.	Opportunity to make a lot of money
		Work/Life balance could be better	Opportunities for growth and advancement
opportunity	work life balance	opportunity for growth, financial compensation, location	work-life balance, job security, culture, tuition reimbursement
		Opportunity for advancement and lack of growth	Family oriented, balanced life, encourage learning
work life balance	management	Long hours and not enough work life balance	Great people and management
		Work / life balance quite difficult	Excellent management at all levels
management	work life balance	Sometimes management is unorganized	The work/Life balance is good, they understand that you have many things going on with your life
		management problems at the middle levels	culture, benefits, compensation, work life balance

4.2 Frequent pro-con pairings

In addition to the 4 pairs constituting the 2 tradeoffs, 8 more pairs have been identified as significantly frequent (see Table 2). None of them constitutes a significantly frequent pro-pro pair except for con: work life balance & pro: culture. Examples can be found in table 4.

Table 2. Summary of frequent pro-con pairs.

con	pro	count	p-val (adj.)	cons pairing	pros pairing
hour	culture	86	0	insignificant	infrequent
	management	132	0	infrequent	infrequent
	pay	177	0	infrequent	infrequent
	training	94	0	infrequent	insignificant
management	benefit	523	0.0001	infrequent	infrequent
pay	work life balance	280	0	infrequent	infrequent
work life balance	culture	77	0,0001	infrequent	frequent
	training	85	0	infrequent	insignificant

Table 4. Examples of reviews illustrating significantly frequent pro-con pairs.

label cons	label pros	cons	pros
work life balance	training	work/life balance. A lot of information to be responsible for	Good benefits and good training
		No work life balance and promise big bonuses but do everything they can to cut bonuses and salary increases at year end.	Gives solid training when starting out.
	culture	Work life balance not the best. Needs work on diversity of employees.	Great culture and good people.
pay	work life balance	worklife balance especially for families.	Great culture and great people.
		pay on the low end	stress is low and work/life balance is good
management	benefit	pay is well below industry average	great work-life balance, knowledgeable people, benefits
		Horrible management. HR management worse.	The benefits are good and not too costly.
hour	training	The management is really horrible.	The benefits are really great.
		Hours, A lot of turnover	Training, no other pros to mention
		Long hours and stressful work environment.	Great training and team environment.

pay	long hours and constant pressure	pay is good hard to beat
	Long hours on the job	Good pay for work done
management	Hours and ongoing changes..	Solid management and great location.
	very long hours raise very small	great work environment management hand off remote work
culture	long hours and over time	like the culture and people
	A lot of hours worked	Great pay and good culture

5 Discussion

5.1 Tradeoffs

The approach identified two tradeoffs: *management vs. work life balance*, and *opportunity vs. work life balance*. A tradeoff between management and work-life balance means that a significant number of people complains about bad management but praises the good work life balance (870, $P < .0001$), while another group of people complains about poor work-life balance but praises the company for good management (118, $P < .0001$). That good work-life balance and good management are not incompatible can be seen from the fact that both labels appear significantly frequently together as a *pro-pro* pair (145, $P < .0001$). This makes intuitive sense: A manager can be good at managing people and care about their work-life balance if circumstances allow for it. However, more often (870, $P < .0001$) employees in finance seem to think that they have bad managers. These managers are potentially less organized, which might result in less pressure at work and thus better work-life balance. For example, one employee wrote: *Cons: Sometimes management is unorganized. Pros: The work/Life balance is good, they understand that you have many things going on with your life* (see the examples in Table 2). Less (but significantly) often, the work-life balance is considered bad, but the management is given credit for doing a good job, for example: *Cons: Long hours and not enough work life balance. Pros: Great people and management.*

The second tradeoff between opportunity and work-life balance showed up in fewer reviews but marks a tradeoff with less compatible labels (significantly infrequent *pro-pro* pairing: 11, $P < .0001$). The tradeoff suggests that working less comes with fewer opportunities for career development in the financial industry (47, $P < .0001$), while having less time outside of work – and therefore probably working more – creates

additional opportunities (54, $P < .0001$). Intuitively, it makes sense and is probably due to time being a scarce resource.

It is striking that work-life balance is involved in both tradeoffs. The concept describes in and of itself a tradeoff evolving around how to allocate one's time. To see whether a solution could be found for *work life balance vs. opportunity* in the sense of Fig. 1, and how to explain *work life balance vs. management*, more analysis would be required, but this is outside of the scope of this paper. For inspiration regarding the former, we could investigate other sectors or use the generalized version of the *algorithm of inventive problem solving* that was designed for dealing with tradeoffs [13].

5.2 Pro-con pairs

The theme underlying the two tradeoffs, i.e., the concern of employees with work-life balance and thus their *time*, is reinforced by the list of significantly frequent pro-con pairs (Table 2). The time-related label, *hours*, appears in the cons section of 4 pairs, and *work life balance* twice in *cons* sections and once in the *pros* section of another pair. 7 out of 8 pairs are thus time related (for a few examples, see Table 4). On the most general level, the explanation seems to be straight forward: employees sell their time to their employer for a salary. If a resource is scarce and different systems exploit the same shared resource, tradeoffs emerge around this resource [14]. While each of the conflicts would deserve further investigation (especially the less obvious ones such as *con: work life balance & pro: culture* (77, $P .0001$), or *con: management & pro: benefits* (85, $P < .0001$), it is outside of the scope of this paper.

5.3 Possible future improvements on the approach

Four main aspects stand out as potential improvements for the approach. First, the classifier to assign labels to the comments requires further improvement. Currently, it is based on a model that has not been fine-tuned for the task of classifying free text on Glassdoor in terms of the identified labels. Second, it is desirable to use more than 13 labels to identify less obvious combinations. Third, it might be worthwhile to do a comparison between sectors. Fourth, the scope of applicability of the approach could be broadened by making it independent of the pros-cons structure that is immanent to reviews on Glassdoor and Indeed. This might be achievable by adding a layer of sentiment analysis. Currently, the approach piggybacks on the pre-categorization of the reviews themselves.

6 Acknowledgment

I am grateful to Glassdoor for making this research possible by providing access to their proprietary data, to Marco Meyer for the discussions of this topic and his advice, to Georg Müller-Christ and his colloquium at the university of Bremen for the feedback and guidance, and to Isabel Estevez for proof-reading the paper.

7 References

1. Glassdoor. <https://www.glassdoor.com>, last accessed 2022/04/29.
2. Indeed. <https://www.indeed.com/>, last accessed 2022/04/29.
3. Altshuller, G. S. *Creativity as an exact science: the theory of the solution of inventive problems*. (Gordon and Breach Science Publishers, 1984).
4. Schumpeter, J. A. *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*. (Porcupine Press, 1939).
5. Tradeoff-detection. <https://github.com/datamaunz/tradeoff-detection>, last accessed 2022/04/29.
6. Guo, F., Gallagher, C. M., Sun, T., Tavoosi, S. & Min, H. Smarter people analytics with organizational text data: Demonstrations using classic and advanced NLP models. *Human Resource Management Journal*, (2021).
7. Luo, N., Zhou, Y. & Shon, J. Employee Satisfaction and Corporate Performance: Mining Employee Reviews on Glassdoor.com. 16 (2016).
8. Smith, D., Choudhury, P., Chen, G. & Agarwal, R. *Weathering the COVID Storm: The Effect of Employee Engagement on Firm Performance During the COVID Pandemic*. <https://papers.ssrn.com/abstract=3841779> (2021) doi:10.2139/ssrn.3841779.
9. Reimers, N. & Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv:1908.10084 [cs]* (2019).
10. sentence-transformers/paraphrase-distilroberta-base-v1 · Hugging Face. <https://huggingface.co/sentence-transformers/paraphrase-distilroberta-base-v1>, last accessed 2022/04/29.
11. spaCy · Industrial-strength Natural Language Processing in Python. <https://spacy.io/>, last accessed 2022/04/29.
12. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6**, 65–70 (1979).
13. Schollmeyer, J. & Tamuzs, V. Deducing Altshuller’s Laws of Evolution of Technical Systems. in vol. AICT-572 55 (Springer International Publishing, 2019). doi: 10.1007/978-3-030-32497-1_6.
14. Schollmeyer, J. & Tamuzs, V. Discovery on Purpose? Toward the Unification of Paradigm Theory and the Theory of Inventive Problem Solving (TRIZ). in *Automated Invention for Smart Industries* (eds. Cavallucci, D., De Guio, R. & Koziółek, S.) 94–109 (Springer International Publishing, 2018). doi:10.1007/978-3-030-02456-7_9.