Global Journal of Engineering Science and Research Management

A SURVEY OF BUSINESS PROCESS SIMILARITY: METRICS AND TECHNIQUES

Rodrigo Costa dos Santos*, Geraldo Bonorino Xexeo
* Programa de Engenharia de Sistemas e Computação, COPPE/ PESC/UFRJ - Brazil
Departamento de Ciência da Computação, IM/UFRJ - Brazil

DOI: 10.5281/zenodo.1143975

KEYWORDS: business process, similarity metrics, survey.

ABSTRACT
This work performs a literature review on the topic of business process similarity. An extensive study was conducted of several indexed academic databases, with the scope of the review being studies of the main topics published in journals and conference proceedings. Additionally, an analysis was done of the papers that have influenced the work of others, through correlation of citations. It can be concluded that the subject began to be explored more intensely in 2007 and, interestingly, there was a surge again in 2012. Finally, the main contribution of this research was a classification of similarity metrics found in the papers, which can be divided into three main groups: structural, textual, and transactional.

INTRODUCTION
Currently, researchers and practitioners are describing organizations in terms of processes rather than functional hierarchies. An organization may be viewed as a chain of interrelated processes that are designed to achieve certain organizational goals.

Additionally, organizations increasingly need to reduce costs and rearrange activities to remain competitive. Thus, companies need tools for planning, management, and control, in order to remain competitive. Restructuring and actions for organizational change are constantly used to increase efficiency for companies' business processes.

Therefore, knowing their internal business processes has become an obligation for companies who want to improve their performance. Business processes and associated activities are the means by which the company adds value to the products and services it provides.

For large companies that have branch locations around the world, or form conglomerates or holding companies, keeping the processes running similarly at all levels becomes a more difficult task. To make this kind of comparison of processes, researchers have focused on studying the similarity between business processes, regardless of the way they are mapped or documented.

Therefore, in order to assemble and study the theories already developed on the topic of business process similarity, this paper intends to conduct a comprehensive review of the literature in an attempt to find what has been written and presented over time regarding the similarity metrics and techniques that have been developed. In total, 35 articles were analyzed [1–35].

At the beginning, some concepts about business processes are reviewed, then the methodology used to perform the survey, the database used for the research, and a comparison of results are shown. Finally, a conclusion is made about the classification of metrics.

REVIEW
The literature review was conducted by exploring the databases of the ACM Digital Library, IEEE Xplore, and the ISI Web of Knowledge, which will herein be referred to as ACM, IEEE, and ISI, respectively. A similar search was performed in the three databases using the terms "business process" and "similarity" as filters for the "title" and "abstract" fields, and without a filter for the year of publication for articles published in journals and conference proceedings up until 2012. The search was done in February 2013.

The number of items returned for each query was: 23 papers for the ACM, 42 for IEEE, and 27 for the ISI. Of the 92 items returned in total, 2 were repeated; that is, they were duplicated. After deleting the redundancies, 90 unique papers remained, as can be seen in Table 1.

Table 1: Amount of papers retrieved

Database	Number of papers retrieved	Duplicated papers removed	Papers considered as relevant	% of relevant papers
ACM	23	21	10	48%
IEEE	42	42	14	33%
ISI	27	27	11	41%
Total	**92**	**90**	**35**	**-**

All 90 papers were analyzed in order to verify their relevance for this research; that is, if the papers contributed to the question of similarity and comparison of business processes. Thus, we read the abstracts, results, and conclusions, of all 90 papers.

After the reading and analysis, only 35 papers were considered to be making a contribution to the topic of business process similarity. These 35 papers constituted the theoretical basis for this research.

As a first analysis, we can highlight the dispersion of the 35 articles according to the year in which they were published. This result can be visualized through the graph in Figure 1.

Since the search had no restriction regarding the initial date, the result of Figure 1 shows that the subject matter is current — the first relevant article was published in 2004 and there was a surge in publications beginning in 2007.

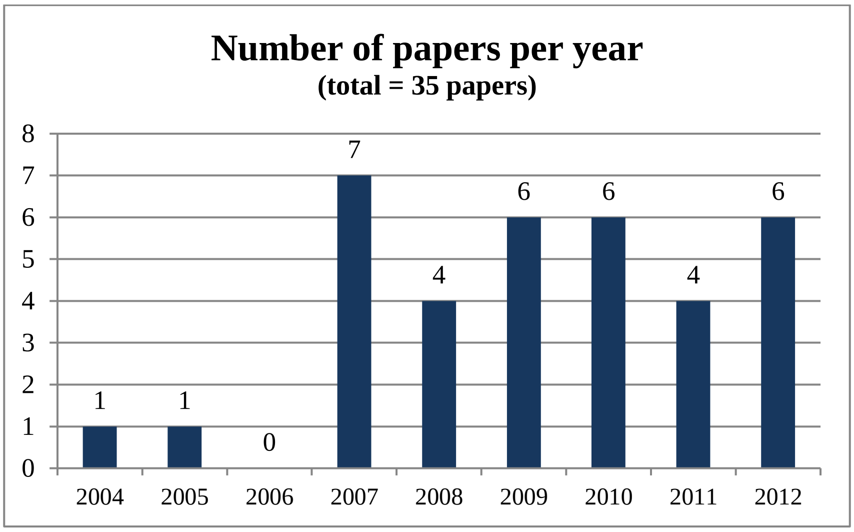


Figure 1: Graphical dispersion of articles by year of publication

It can also be seen that some articles were very similar to each other, due to having the same author or the same group of authors. The works that presented minor changes, or were published in different journals or on different indexing bases are: [8, 11, 16]; [12, 19]; [13, 20]; [14, 25]; [21, 34]; and [22, 24].

The details and information are in Table 2. All papers were read in full and the information necessary to conduct the study was extracted as follows:
a) Goal/motivation - what was the motivation of the author for exploring the topic of business process similarity?
b) Type of notation for the process - what type of notation was used in the paper?
c) Similarity metrics - what metrics were presented?
d) Business process database used for validation - identifies whether or not there was validation in the article and, if so, which business process database was used.

 ISSN 2349-4506
Impact Factor: 3.779

Global Journal of Engineering Science and Research Management

Based on the information in Table 2, it was possible to identify some important findings. The first finding was related to the motivation for the papers on the subject of similarity, which is summarized and presented in Table 3.

Table 2: Analysis of relevant papers

Ref.	a. Syntactic	b. Semantic	a. Comparison of elements	b. Vector and clustering	c. Transition adjacency relation (TAR)	d. Graph edit-distance	e. Task context	a. Traces and Logs analysis
[1]	-	-	✓	-	-	-	-	-
[2]	-	-	✓	-	-	-	-	-
[3]	-	✓	-	-	-	-	-	-
[4]	✓	✓	-	-	-	-	-	-
[5]	-	-	✓	-	-	-	-	-
[6]	-	-	-	-	-	✓	-	-
[7]	✓	✓	-	-	-	-	✓	-
[8, 11, 16]	✓	✓	-	-	-	-	-	-
[9]	-	-	✓	-	-	-	-	-
[10]	-	-	-	-	-	✓	-	-
[12, 19]	-	-	-	✓	-	-	-	-
[13, 20]	-	✓	-	-	-	-	-	-
[14, 25]	-	-	-	-	✓	-	-	-
[15]	✓	-	-	-	-	✓	-	-
[17]	-	-	-	-	-	✓	-	-
[18]	-	-	✓	-	-	-	-	-
[21, 34]	✓	✓	-	-	-	✓	-	-
[22, 24]	-	-	✓	-	-	-	-	-
[23]	-	-	✓	-	-	-	-	-
[26]	-	-	-	-	-	-	-	✓
[27]	-	-	-	-	-	-	-	✓
[28]	✓	-	-	-	-	-	-	-
[29]	✓	✓	-	-	-	✓	✓	-
[30]	-	-	-	-	✓	-	✓	-
[31]	-	-	-	-	-	✓	-	-
[32]	✓	-	-	-	-	-	-	-
[33]	✓	✓	-	✓	-	-	-	-
[35]	-	✓	-	-	-	✓	-	-

The two main lines of research in this area are: (1) the development of algorithms and similarity metrics for the data of a business process input, in order to find the one most similar to a process repository; and (2) the development of similarity metrics that, given two distinct business processes, will identify their similarity, typically in a range from 0 to 1. Together, these two motivations were present in 20 of the 35 relevant articles surveyed.

Another motivation that also stood out was the use of the similarity between business processes to identify the lack of alignment between the actual processes of the company and those that are implemented in the Enterprise Resource Planning (ERP) system. In this sense, one can see that the ERP was implemented and customized in the companies in the best possible way [2, 8, 11, 16].

A possible application was also presented in [25, 29], using the similarity between business process models to identify similarities in the context of acquisitions and mergers of companies with several geographically spread branches. With this tool, process analysts could identify common or similar processes between companies or subsidiaries in order to examine their overlaps and redundancies, and also to identify areas that could be consolidated.

Paper [10] highlights the possibility of minimizing the efforts to transform a process into another one — the aim being to support project redesign of business processes.

Table 1: Motivation of the papers analyzed

Purpose of Paper/Motivation	**Quantity**
Search for the most similar process in a repository	12
Compare two related business processes	8
Detect deficiencies between the ERP process and business needs	4
Identify fragments of similar processes	2
Present similarity techniques	1
Compare two similarity techniques	1
Detect variants of processes and facilitate process redesign	1
Graphically display the unification process	1
Identify opportunities for refactoring processes	1
Promote integration between business processes	1
Infer similarity between business processes through analysis of Logs	1
Identify consistencies in workflows	1
Search and merge state machine diagram	1

Among the types of notation used in the papers of the processes studied, there is a predominance of Petri nets, EPC, and BPMN, as shown in the graph of Figure 2. However, to perform similarity calculations and structural comparisons between the processes, the authors transformed these processes into a simpler notation, called "graph" — much of which has definitions of the authors. Notations such as BPMN and EPC are quite complex and have various representation symbols; however, for treating the similarity, the most important items were the tasks, their structural arrangements, and connections.

Global Journal of Engineering Science and Research Management

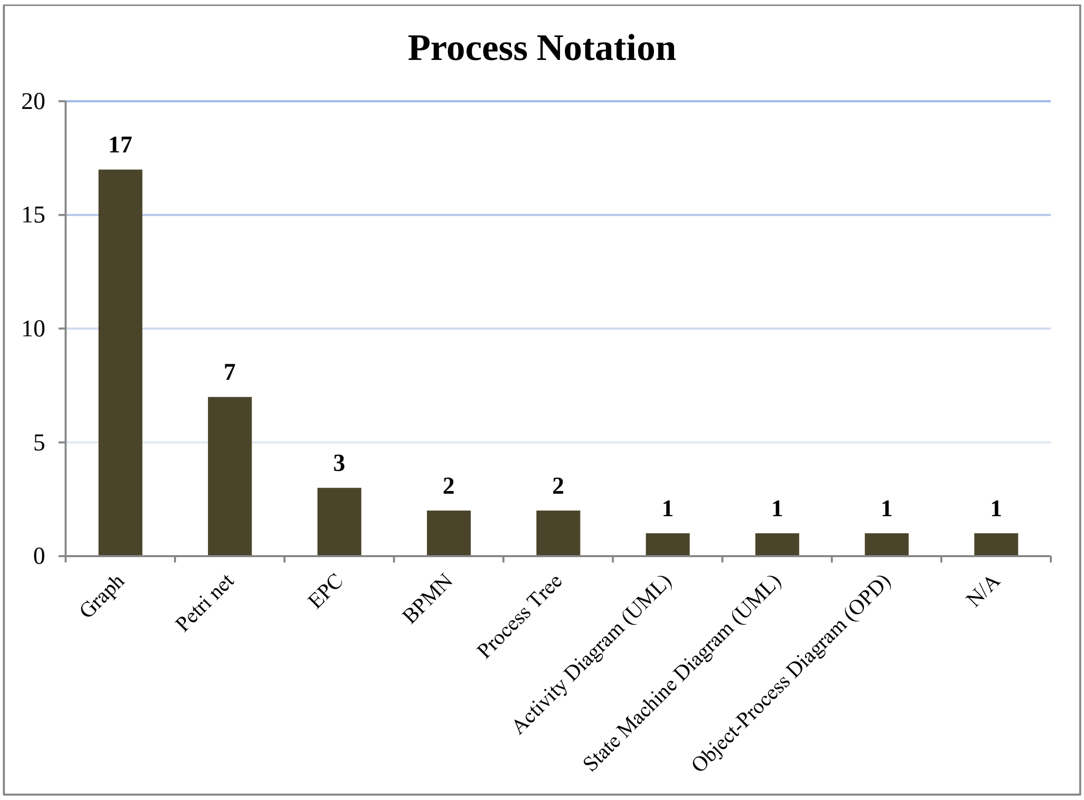
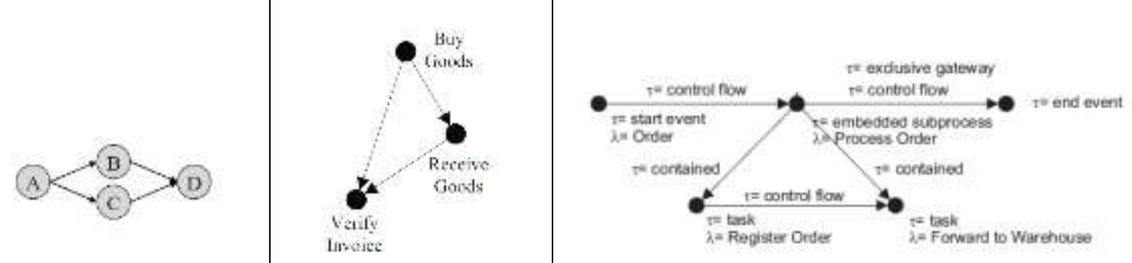


Figure 2: Notation used in the articles

The simplest definition of graphs [9, 12, 19, 17] contains only two elements: tasks and edges, which represent the connections between the tasks, as shown in the graphical representation of Figure 3a. Other definitions include a set of labels [21, 15, 26], as shown in Figure 3b. Finally, other definitions contain four elements: tasks, edges, label, and type of edge and event [29, 33, 31], similar to the standard EPC, as shown in Figure 3c.



(a) ***Based on [9]*** | **(b)** ***Based on [21]*** | **(c)** ***Based on [29]***

Figure 3: Types of graphs presented in the papers analyzed

In order to verify the influence of a work on other works, the correlation between the references of the 35 papers was another relevant analysis. We did correlation mapping between the citations of the 35 papers. We found that 22 papers cited at least one of the remaining articles; 17 papers were cited at least once; 10 papers were listed more than once; and 7 were cited only once — see Table 4.

Global Journal of Engineering Science and Research Management

Of all the papers, there were only 6 that neither made citations, nor were they cited by other papers. This demonstrates quite a strong relationship among the 35 papers surveyed, since each of the other 29 articles (83%) was cited or mentioned at least once.

Table 2: Papers that were cited and frequency of citations

No. of times cited	Paper
14 times	[4]
8 times	[10]
7 times	[5]
6 times	[29]
5 times	[6]
4 times	[15]
twice	[2, 7, 21, 25]
once	[3, 8, 9, 13, 18, 23, 28]

In order to present the citation network among the papers, we designed a Fruchterman-Reingold type graph using Network Workbench[1] where each node represents one paper, and the size and color correspond to the number of times the article is cited. The result can be seen in the graph of Figure 4.

Analyzing Figure 4, we can see that the papers citing only one other article (smaller nodes) and which were not cited themselves, are located at the outer limits. The articles that are cited and also cite other articles are located more towards the center. It can be gathered from this assessment that the 12 papers which only make citations are known as the "influenced works"; that is, they are based on the theory of other research and can extend the knowledge or create new developments from the theory already presented.

Meanwhile, in relation to the 17 papers at the center of the graph, these works can be considered to be influential because some of them have produced theory that was evaluated and found to be valid and subsequently expanded by other authors. Among the articles, we highlight the influential [4], [10], [5], [29], and [6] — in this order, they form the 5 most cited articles, as can be seen in Table 4.

In assessing the chronological order of citation, it can be seen that articles [4], [5] and [6], which were published in 2007, are the pioneers for the topic of business process similarity. Despite these being the first works, other studies have become influential over time. It is worth mentioning article [29], which, although being published in 2011, collected 6 citations between 2011 and 2012.

[1] Available to download at: http://nwb.cns.iu.edu/

Global Journal of Engineering Science and Research Management

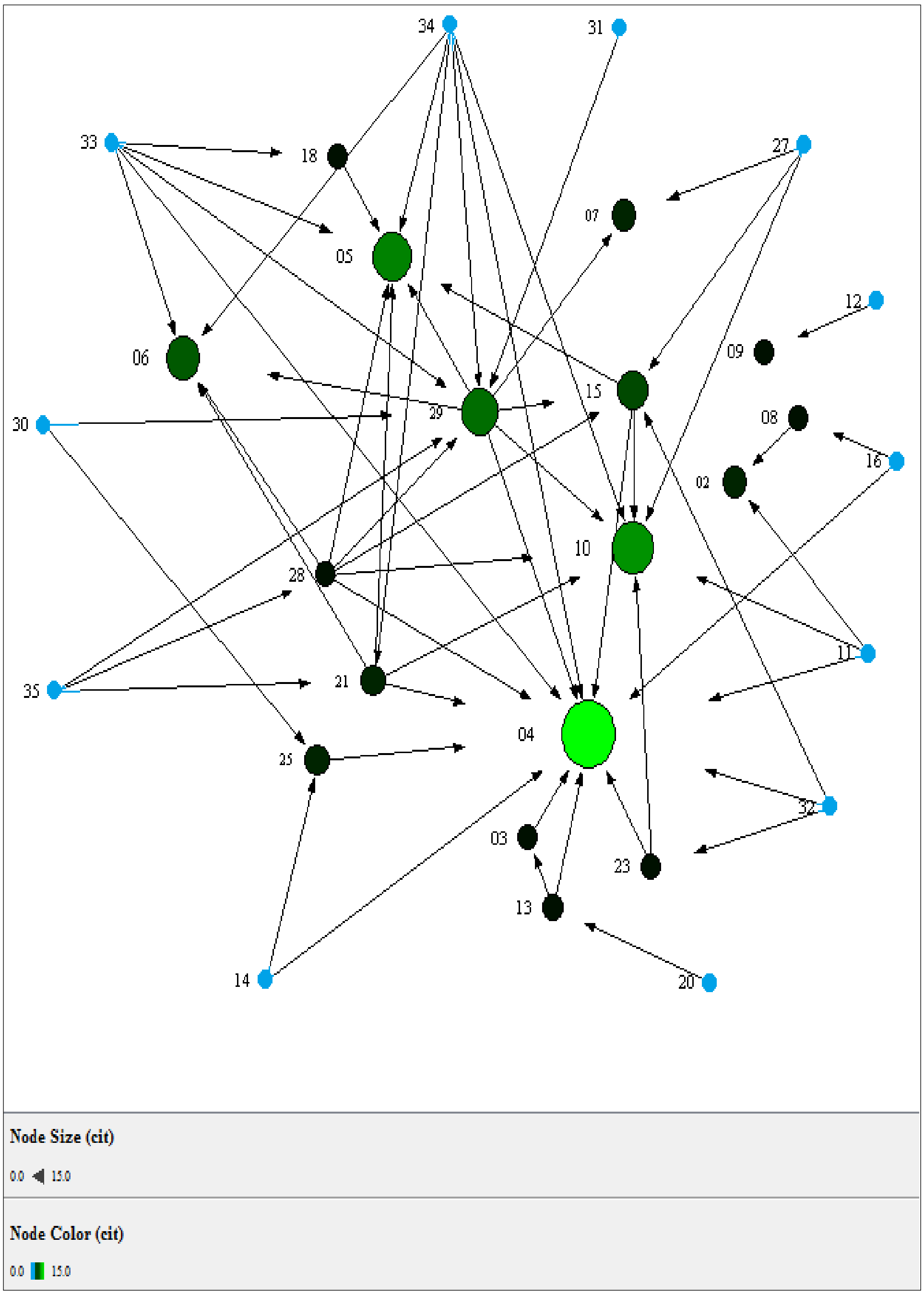


Figure 4: Citation network among the papers

SIMILARITY METRICS

The most important point of this research was to identify the similarity metrics of business processes used and presented in the articles. These similarity metrics can be divided into three main groups: (i) **textual**, based on comparison of the labels of the task, one by one; (ii) **structural**, based on comparison of the structure of the graph of the processes as a whole; and (iii) **transactional**, based on the analysis of system logs and traces.

Each of the three groups presented is composed of one or up to several different metrics, which is addressed by the authors of the papers. Each author presented a combination of metrics that were analyzed and classified. The metrics are summarized and presented in Figure 5.

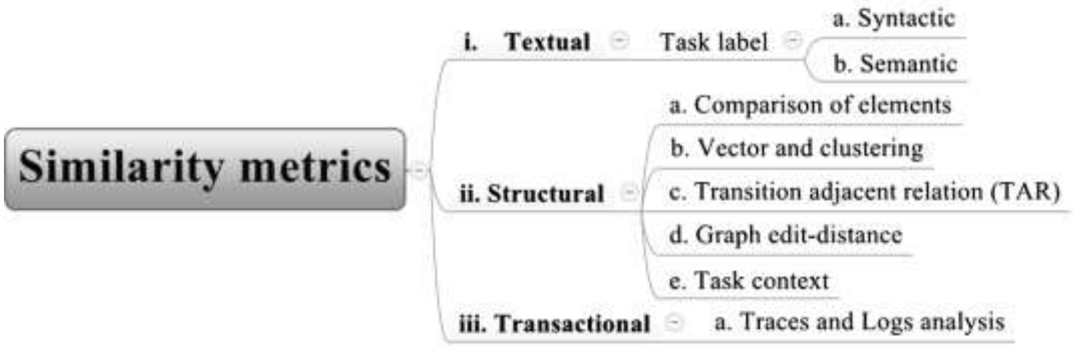


Figure 5: Mapping of the metrics used in the papers analyzed

Textual
The textual metrics aim to compare the descriptions of the tasks of business processes. These metrics can be classified as syntactic or semantic.

Syntactic
The syntactic metric was applied in 12 of the 35 papers surveyed [4, 7, 8, 11, 15, 16, 21, 28, 29, 32, 33, 34]. The most widely used type of calculation, present in 9 of the 11 articles, the edit distance technique was proposed by Levenshtein [36]. The edit distance technique measures the difference between two strings. The Levenshtein edit distance between two words is equal to the minimum number of edits needed to transform one word into another through the insertion, deletion, or substitution of a single character. The Levenshtein distance can be determined by the cost to compute the modify operations between two strings.

For example, the edit distance between the strings "informer" and "information" is equal to 5, because 2 substitutions and 3 insertions are needed to transform "informer" to "information". The proposed syntactic similarity measure, which returns degree of similarity values between 0 and 1, is shown in equation 1.

$$\mathrm{Sim}(s,t) = 1.0 - \frac{\mathrm{ed}(s,t)}{\max(|s|,|t|)} \quad (1)$$

The function ed(s,t) denotes the minimum number of operations needed to transform s into t or vice versa. The max(|s|, |t|) represents the greatest length between the two strings. Applying the equation 2 to the example shown, sim(s, t) = 1.0 - (5/11) ≈ 0.45.

The authors of [33] used the Jaccard similarity coefficient [37] as an alternative for comparing tasks' labels. The Jaccard coefficient measures similarity between sets of strings and is defined as the size of the intersection divided by the size of the union of the sample sets.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

Using the example "deliver the book" and "book delivery", by applying the Jaccard similarity coefficient, we have four different words as the union and one word as the intersection, and they are assigned a similarity value of (1/4) = 0.25. On the other hand, if using stop word removal and lemmatizing, the result is (2/2) = 1. The challenge is to find a good metric; however, word preprocessing combined with other metrics can be useful.

Besides these two metrics, as presented in [7] and [32], n-gram was also used to calculate similarity. N-gram is a sequence of n letters, where n is the index. The n-grams with n = 1, n = 2, and n = 3 are called monogram, bigram, and trigram, respectively. This technique consists of examining each n-gram of a word entry and searching in a table of valid n-grams [38]. The work in [7] used a similarity algorithm with trigrams and [32] used bigrams.
Semantic

Global Journal of Engineering Science and Research Management

The semantics metric was used in 13 of the 35 papers surveyed [3, 5, 7, 8, 11, 13, 16, 20, 21, 29, 33, 34, 35]. The comparison of synonyms using WordNet was unanimous. WordNet is a database of English language knowledge containing the meaning of words, their context synonyms, and homonyms, among other attributes [39]. As in [39], the possibility of using hypernyms and hyponymy to improve the evaluation of semantic similarity is presented; however, this was not tested.

Structural

The structural metrics aim to compare the structure of the process, comparing the elements of graphs, their types, inputs and outputs, regardless of job descriptions. Several studies describe the activities generically; for example, as "A" or "B", and use this for comparison of the activity to determine if it is identical or not; that is, 0 or 1. Therefore, only the structure of the graph is taken into account in structural metrics.

Comparison of elements

In [1, 2, 5, 9, 18, 22, 23, 24], a similarity metric was developed to compare the elements that make up the process. In [5, 22, 24], the structural similarity is calculated based on similarity and dissimilarity between the structural elements of process variants. In [23], the equivalence between fragments of process models is calculated based on the comparison of the types of elements — a comparison of tasks and their execution order is added.

Vectorization and clustering

The authors of [12, 19, 33] show the similarity metric of comparing processes using vectorization and clustering techniques.

Vectorization is presented in [12, 19]. The authors use two types of vector models to express the structure of a business process: one for activities and one for the transitions. Vectors are used to calculate the co-occurrence activities and their dependencies, when there are two cases. The cosine coefficient is adopted to measure the similarity between two process models. The highest cosine coefficient between two vectors is the most common element with the highest value.

In [33], after application of the semantics metric to the terms, the author applied a clustering technique between fragments of the process. Information on the structural similarity and its association with the fragments allows the creation of pairwise comparison processes.

Transition Adjacency Relations (TAR)

Works [14, 25] introduced a term known as transition adjacency relations (TAR). The TAR set describes order transitions appearing in all the possible sequences within the process (one directly followed by the other). After this is done, it is possible to build a matrix with all transitions between process tasks and then compare with the matrix of another process. With the TAR concept it is possible to generate a metric for distances between two processes. In [30], the TAR concept was also explored and this generated some experiences for this metric.

Graph edit distance

The edit-distance metric of a graph is the most commonly used structural metric among the searched items [6, 9, 10, 15, 17, 21, 29, 6, 34, 31, 35], being present in 10 of the 35 cited papers. The notion of edit distance for the graph is the same as the textual syntax; that is, the Levenshtein edit distance [36], so that instead of characters, the insertion, deletion, or substitution operations are done on the nodes of the graph (tasks or events).

The disadvantage of this metric is that most studies assume the tasks and identical events through letters or generic identifications. Only [15, 21, 34, 29] used a combination of textual and structural metrics.

Task context

The papers [7, 29, 30] defined a metric to determine the contextual similarity between elements of a business process model. To apply the contextual metric, it is necessary to establish the equivalence between elements in their input contexts and also the equivalence between the elements of contextual output.

This metric calculates a degree of similarity for each pair of states or tasks, aggregating the degrees of similarity between the immediate neighbors. It is understood by neighboring states and task successors and predecessors. The similarity algorithm between neighbors can vary, according to the survey.

Global Journal of Engineering Science and Research Management

According to [30], the similarity between two nodes depends not only on the nodes but also on their context.

Transactional

Traces and Logs analysis

This way of analyzing the similarity is quite specific and consists of the observation records that represent the executions of transitions, as recorded by information systems — it is also known as a System Log. Papers [26, 27] present this technique. First, there is a formal definition of the event logs of a system, then tasks and events can be extracted.

CONCLUSION

In this work we performed a literature review on the topic of business process similarity. Mapping and managing business processes remains a good strategy for finding bottlenecks, single points of failure, and excesses and shortages of manpower throughout the production chain of a company.

Knowing and then trying to map your processes, the first difficulty arises; that is, the choice of notation to be used. There are several on the market, amongst them BPMN, EPC, UML, and Petri Nets, each with its own peculiarities. This number of notations creates the first impediment for researching processes, their comparisons, and benchmarking. Researchers utilize a graph structure which is a simplification of the current notations and acts as if they were a universal conversion for representing processes.

Another point that can be concluded is that the main goal pursued by researchers is to determine the most similar search and recovery process in a repository. To fulfill this task, it is necessary to make use of similarity metrics between business processes.

Several metrics have been presented in the publications on the subject. After a detailed study, it is possible to arrive at a classification of metrics for business process similarity, which represents the main contribution of this research.

Among the different metrics are some focused on the structure of the process as a whole; for example, the analysis of inputs and outputs, number of tasks, arcs, and decisions. Other metrics note the description of the tasks and try to infer syntactic and semantic similarity, especially with the use of synonyms and the elimination of stop words to enhance the search results. Some less common metrics, but ones which deserve to be presented, are those related to reading traces and system logs.

The present study also shows that it is relevant to make a correlation between the date of publication of the articles and the analysis of which authors influenced the work of others. It can be concluded that the subject began to be explored more intensely in 2007 and there was a surge again in 2012.

Attention is called to the fact that so many metrics are presented over time, but few studies have comprehensively validated these metrics using reliable and consistent repositories. A recommendation for future work is to conduct research that validates the metrics consolidated in this work by combining this set of metrics to arrive at a conclusion about their effectiveness.

REFERENCES

1. G. GROSSMANN, M. SCHREFL, M. STUMPTNER. (2004) Classification of Business Process Correspondences and Associated Integration Operators. in Proc. ER (Workshops), pp.653-666.
2. P. SOFFER, B. GOLANY, D. DORI. (2005) Aligning an ERP system with enterprise requirements: An object-process based approach. Presented at Computers in Industry, pp.639-662.
3. A. KOSCHMIDER A. OBERWEIS. (2007) How to detect semantic business process model variants?. in Proc. SAC, pp.1263-1264.
4. M. EHRIG, A. KOSCHMIDER, A. OBERWEIS. (2007) Measuring Similarity between Semantic Business Process Models. in Proc. APCCM, pp.71-80.
5. R. LU AND S.W. SADIQ. (2007) On the Discovery of Preferred Work Practice Through Business Process Variants. in Proc. ER, pp.165-180.

6. M. MINOR, A. TARTAKOVSKI, R. BERGMANN. (2007) Representation and Structure-Based Similarity Assessment for Agile Workflows. in Proc. ICCBR, pp.224-238.
7. S. NEJATI, M. SABETZADEH, M. CHECHIK, S.M. EASTERBROOK, P. ZAVE. (2007) Matching and Merging of Statecharts Specifications. in Proc. ICSE, pp.54-64.
8. J. GAO, L. ZHANG, W. JIANG. (2007) Procuring Requirements for ERP Software Based on Semantic Similarity. in Proc. ICSC, pp.61-70.
9. J. BAE, L. LIU, J. CAVERLEE, L. ZHANG, H. BAE. (2007) Development of Distance Measures for Process Mining, Discovery and Integration. presented at Int. J. Web Service Res., pp.1-17.
10. C. LI, M. REICHERT, A. WOMBACHER. (2008) On Measuring Process Model Similarity Based on High-Level Change Operations. in Proc. ER, pp.248-264.
11. J. GAO, L. ZHANG. (2008) Detecting Gaps between ERP Software and Organizational Needs: A Semantic Similarity Based Approach. in Proc. Semantic Computing and Systems, pp.21-26.
12. J. JUNG, J. BAE, L. LIU. (2008) Hierarchical Business Process Clustering. in Proc. IEEE SCC (2), pp.613-616.
13. A. AWAD, A. POLYVYANYY, M. WESKE. (2008) Semantic Querying of Business Process Models. in Proc. EDOC, pp.85-94.
14. H. ZHA, J. WANG, L. WEN, C. WANG. (2009) A label-free similarity measure between workflow nets. in Proc. APSCC, pp.463-469.
15. R.M. DIJKMAN, M. DUMAS, L. GARCÍA-BAÑUELOS, R. KÄÄRIK. (2009) Aligning Business Process Models. in Proc. EDOC, pp.45-53.
16. J. GAO, L. ZHANG. (2009) On Measuring Semantic Similarity of Business Process Models. in Proc. IESA, pp.289-293.
17. J.ZHU, H.K. PUNG. (2009) Process Matching: A Structural Approach for Business Process Search. in Proc. COMPUTATIONWORLD, pp.227-232.
18. K. ANDREWS, M. WOHLFAHRT, G. WURZINGER. (2009) Visual Graph Comparison. in Proc. IV, pp.62-67.
19. J. JUNG, J. BAE, L. LIU. (2008) Hierarchical Business Process Clustering. in Proc. IEEE SCC (2), pp.613-616.
20. S. SAKR, A. AWAD. (2010) A framework for querying graph-based business process models. in Proc. WWW, pp.1297-1300.
21. Z. YAN, R.M. DIJKMAN, P. GREFEN. (2010) Fast Business Process Similarity Search with Feature-Based Similarity Estimation. in Proc. OTM Conferences (1), pp.60-77.
22. N.M.B. MAHMOD, S.B.A. RADZI, An approach to analyse similarity of business process variants. In Proc. Progress in Informatics and Computing (PIC), 2010,pp.640-644.
23. C. GERTH, M. LUCKEY, J.M. KÜSTER, G. ENGELS. (2010) Detection of Semantically Equivalent Fragments for Business Process Model Change Management. In Proc. IEEE SCC, pp.57-64.
24. N.M. MAHMOD, W.Y. CHIEW. (2010) Structural similarity of business process variants. In Proc. ICOS, pp.17-22.
25. H. ZHA, J. WANG, L. WEN, C. WANG, J. SUN. (2010) A workflow net similarity measure based on transition adjacency relations. In presented at Computers in Industry, pp.463-471.
26. Y. LI. (2011) Discovery of role-based organization structure model in BPMS. In Proc. FSKD, pp.2022-2025.
27. M. WEIDLICH, A. POLYVYANYY, J. MENDLING, M. WESKE. (2011) Causal Behavioural Profiles - Efficient Computation, Applications, and Evaluation. Presented at Fundam. Inform., pp.399-435.
28. R.M. DIJKMAN, B. GFELLER, J.M. KÜSTER, H. VÖLZER. (2011) Identifying refactoring opportunities in process model repositories. Presented at Information & Software Technology, pp.937-948.
29. R.M. DIJKMAN, M. DUMAS, B.F.V. DONGEN, R. KÄÄRIK, J. MENDLING. (2011) Similarity of business process models: Metrics and evaluation. Presented at Inf. Syst., pp.498-516.
30. J. WANG, S. TAN, L. WEN, R.K. WONG, Q. GUO. (2012) An empirical evaluation of process mining algorithms based on structural and behavioral similarities. In Proc. SAC, pp.211-213.
31. B. CAO, J. YIN, S. DENG, D. WANG, Z. WU. (2012) Graph-based workflow recommendation: on improving business process modeling. In Proc. CIKM, pp.1527-1531.
32. M.C. BRANCO, J. TROYA, K. CZARNECKI, J.M. KÜSTER, H. VÖLZER (2012) Matching Business Process Workflows across Abstraction Levels. In Proc. MoDELS, pp.626-641.

Global Journal of Engineering Science and Research Management
33. M. NIEMANN, M. SIEBENHAAR, S. SCHULTE, R. STEINMETZ. (2012) Comparison and retrieval of process models using related cluster pairs. Presented at Computers in Industry, pp.168-180.
34. Z. YAN, R.M. DIJKMAN, P.W.P.J. GREFEN (2012) Fast business process similarity search. Presented at Distributed and Parallel Databases, pp.105-144.
35. R.M. DIJKMAN, M.L. ROSA, H.A. REIJERS. (2012) Managing large collections of business process models - Current techniques and challenges. Presented at Computers in Industry, pp.91-97.
36. JACCARD, P. (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. In: Bulletin del la Société Vaudoise des Sciences Naturelles 37, pp. 547-579.
37. KONDRAK, GRZEGORZ. (2005) N-Gram Similarity and Distance. In: SPIRE 2005. Springer-Verlag Berlin Heidelberg, pp. 115–126.