

Regulation Relatedness Map Creation Method with Latent Semantic Analysis

Mehmet Murat Huyut^{1,*}, Batuhan Kocaoğlu² and Ünzile Meram³

¹Kuveyt Türk Participation Bank, R&D Center, Kocaeli, 41420, Turkey

²Piri Reis University, MIS Department, Istanbul, 34940, Turkey

³Management Engineering Department, Bahcesehir University, Istanbul, 34353, Turkey

*Corresponding Author: Mehmet Murat Huyut. Email: mhuyut@gmail.com

Received: 08 October 2021; Accepted: 25 January 2022

Abstract: Regulatory authorities create a lot of legislation that must be followed. These create complex compliance requirements and time-consuming processes to find regulatory non-compliance. While the regulations establish rules in the relevant areas, recommendations and best practices for compliance are not generally mentioned. Best practices are often used to find a solution to this problem. There are numerous governance, management, and security frameworks in Information Technology (IT) area to guide businesses to run their processes at a much more mature level. Best practice maps can be used to map another best practice, and users can adapt themselves by the help of this relation maps. These maps are created generally by an expert judgment or top-down relationship analysis. These methods are subjective and easily create inconsistencies. In order to have an objective and statistical relationships map, we propose a Latent Semantic Analysis (LSA) based model to generate a specific relatedness correlation map. We created a relatedness map of a banking regulation to a best practice. We analyzed 224 statements of this regulation in relation to Control Objectives for Information Technologies (Cobit) 2019's 1202 activities. Furthermore, we support our LSA results with MCDM analysis methods; Fuzzy Analytics Hierarchy Process (FAHP) to prioritize our criteria and, WASPAS (Weighted Aggregated Sum Product Assessment Method) to compare similarity results of regulation and Cobit activity pairs. Instead of the subjective methods for mapping best practices and regulations, this study suggests creating relatedness maps supported by the objectivity of LSA.

Keywords: It compliance; LSA; it regulation; Cobit; fuzzy AHP; WASPAS

1 Introduction

In the IT world creating standard processes or techniques are crucial. Network devices use protocols to connect or server client architecture uses almost same rules to communicate. Rule creation is essential to continue IT services. On process side, some institutes create best practices to govern



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IT better. Also, there are some best practices that are more technical issues in. Another dimension is regulations [1]. Every country has its own regulations to create rules about IT, even some sectors private regulations define more specific rules to obey, but best practices are not sector-specific and creates general detail practices and recommendations. Also, best practices and regulation relatedness were never worked by the regulator. Regulatory agencies use best practices to address some statements but in general, there is no defined relatedness between them. Best practices compliance direct relatedness is too little or any. According to every country's own regulation, control objectives, control practices importance levels are changing. Although some best practices are mapped with other best practices to easily apply, its map with countries' special regulation is not defined. Best practices to other standards mapped with each other but this mapping generally expert's subjective views and a top-down approach, they do not match every practice or a statement they generally match master domains.

Our selected regulation is a basis for banks in Turkey for whole IT processes, it regulates the IT areas and create rules to obey. Regulators mention their rule in their regulations but they do not explain how to be compliant. To be compliant in a specific article or an area banks uses best practices. It can be ISO, NIST or Cobit to find related activities. Finding related activities, practices or a domain is a hard process. Best practices domain, subdomain or practices names and creation methods are different, regulations are complicated. To create a semantic relatedness map can improve this process.

To find semantic similarities in texts, there are 5 main methods as; string-based, character-based, corpus-based, knowledge-based, and hybrid similarity measurements methods. [2] Corpus-based methods are Hyperspace Analogue to Language (HAL), Latent Semantic Analysis (LSA), Explicit Semantic Analysis (ESA), Cross-Language Explicit Semantic Analysis (CLESA), and Pointwise Mutual Information-Information Retrieval (PMI-IR). These methods are based on semantical analysis with a predefined corpus word-to-word relation corpus. There are discussions about which method is the best way to find paragraph similarity [3] and also, different results found in the literature [4], Latent Semantic Analysis LSA is an appropriate method for long paragraph similarity finding. Reference [5] We suggest LSA can be used to map regulatory compliance requirements to a selected best practice Cobit 2019 activities. Cobit is an IT and business framework for the governance and management of enterprise IT which is created by ISACA (Information Systems Audit and Control Association) We use this semantic method to find meaning and define every regulation statements' relationship percentage of every Cobit activities to create a relatedness map.

LSA is a type of neural language processing subset of automatic neural language processing based on predefined word-to-word association maps called corpus. The underlying idea is that the sum of information about all word contexts, in which a given word appears and does not appear, largely determines the similarity of the word meaning and word set. This provides a set of mutual constraints and with the help of this method, text similarities can be defined after creating an analyzable data set.

In this study, one to many LSA method was used to compare one regulation statement to many Cobit activities. After defining every regulation statements' relation percentage of every Cobit practices relatedness percentage, we create 224*1202 relation matrices. After that, we analyzed our LSA-based data set with FAHP and WASPAS to prioritize our criteria and rank comparison pairs. Comparison pairs are one of randomly selected regulation statement and LSA result that most or least meaningful pairs selected from Cobit activities.

Finding the relatedness of regulation-is requiring intense effort for governing bodies or compliance officers. Gap analyses or finding related compliance requirements is generally hard. There are not enough relatedness maps created. Only best-known best practices to another best practice relatedness maps or relation matrices made by generally subjective methods. E.g. Cobit to ISO 27001.

Regulations to another best practice relatedness mapping has never worked before, because of the creation process of relation map difficult. NLP (Natural Language Processing) and semantic analysis techniques are being developed and using in many areas. However, these methods are not applied to define relation of regulations or best practices before. Our works are unique in this aspect. Not only in this implementation but also, we used our LSA results with FAHP + WASPAS method to compare created LSA matrices. This strengthens the accuracy of this study and offers a new method to map regulation to best practices or both.

2 Literature Review

In some sectors risks and compliances are changing fast, and it is difficult to be compliant. Businesses want to have quality processes by adapting to best practices. In doing so, there are regulations issued by the regulators that businesses have to comply with. Many of the practices included in the provisions of the legislation are included in international good practices, that is, in best practices. While businesses comply with the legislation, they also want to comply with international practices. Therefore, the compatibility of international rules such as Cobit 2019 with local legislation is always one of the issues investigated by businesses. In the study, [6] included in the literature, Cobit 2019 and ISO 27001 (International Organization for Standardization) regulations were evaluated simultaneously.

Cobit and NIST Cyber Security Framework or other references maps were created manually by Nicole Keller in 2018, and also Glenfis, mapped Cobit 2019 with ITIL v4 relatedness map with only their master domain name similarity [7]. Another way to map standards is to use a unified definition language and semantic analysis to find the relatedness of all the clauses, especially the ArchiMate unified markup language (UML) used in some researches [8]. ArchiMate samples for ISO standards relatedness with Cobit. In the literature, there are some specific domain-to-domain relatedness map created too. Not a whole standard relatedness map but a partial status to research. Cobit DSS02 to ISO 33072 map made with the help of ArchiMate UML in another research [9].

Cobit and a specific IT regulation relatedness has not been worked, Although Cobit vs. ISO standards or ITIL like best practices relation papers created, there is not a sample directly to a specific regulation to Cobit relatedness. In a paper, researchers measure semantic similarities across EU GDPR and cloud privacy policies [10]. Natural language understanding is a trending topic among researchers about finding similarity/relatedness between academic writing or any word knowledge to create scientifically relatedness analysis [11].

Semantic relatedness has several applications in NLP such as word sense disambiguation, paraphrasing, text classification, dimension reduction, etc. [12]. Most semantic relatedness techniques stress the concept that comparable words tend to appear in similar situations and have similar contextual distributions. These approaches frequently provide a distributional semantics (DS) model, which depicts each linguistic word as a vector obtained from contextual information about that phrase in a huge corpus of text or knowledge base [13–16].

To cluster mashup web technology researchers focus on utilizing semantic similarities to guide the mashup clustering process and to find structural and semantic information in mashup profiles. They used to integrate structural similarity and semantic similarity using fuzzy AHP and LDA (Latent Dirichlet Allocation). LDA is a clustering algorithm based on the genetic algorithm again an LSA-like NLP technique [17].

The entire literature on highly correlated maps generated using LSA and fuzzy AHP was reviewed. There is a lack of studies, no comparison of an international regulation such as Cobit 2019 with another regulation with the LSA method has not been found before. In addition, no study has been found in the literature supporting the LSA method with the fuzzy AHP method.

This study aims to reveal a scientific similarity map by using the LSA method by extracting the similarity map of Cobit 2019 and the related regulation. As a result of this analysis, the results of the study were expressed more strongly by using the fuzzy AHP method. Because of all these, the hypothesis of our study can be determined as follows: LSA method can be used in the regulation and best practice mapping and support our LSA result with fuzzy AHP and WASPAS methods used to determine the most important similarity criteria and ranked similarity of our randomly selected Cobit 2019 and regulation statement similarity is aligned with LSA results. Finally, we analyzed that created relatedness map is consistent. Fig. 1. shows our study's main steps from beginning to end.

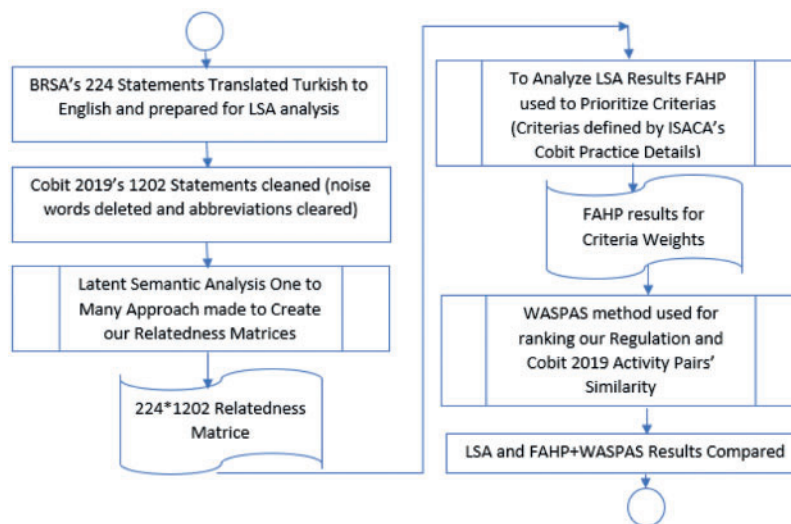


Figure 1: Flowchart of our study

In the third section we expressed all equations which are used in LSA, FAHP and WASPAS, in the fourth section we implemented all methods and represent application results and in the last section, we conclude and give recommendations.

3 Equations and Methods

3.1 LSA Expressions

LSA is a method that puts relevant data from a large document data with a particular query, allowing us to detect similarities between data. The query process has several development routes, including keyword, weighted keyword matching, and vector-based relationships based on word formation [18]. One of these methods, the vector-based approach, has been developed and the single value decomposition (SVD) method has emerged. SVD is a matrix algebra technique that reorients and orders dimensions in a vector space.

LSA extracts information from phrases or phrases that often appear simultaneously in different sentences. If there is more than a one-word group in the sentences in the specified database, the sentence has a semantic or safe meaning [19]. Basically, LSA steps are mentioned below [20]: text preprocessing,

creating a document matrix term, singular value decomposition (SVD) calculation, and calculating the vector value for each document. LSA one-to-many approaches creates SVD and uses cosine similarity.

3.1.1 Collect a Data Set

The database of the text to be processed is determined and separated into documents. Considering that each item in the database is consistent, each item is processed as a separate document. This stage is the stage of transforming the texts in the unstructured database into structured data.

3.1.2 Create a Matrix

At this stage, the structured document and terms matrix are created. In this matrix consisting of rows and columns, it is determined how many times the specified term occurs in the specified database. Thus, the document-term matrix is created. Each row represents a word, while each column represents the resulting word root.

3.1.3 Calculating SVD and Vector Value

The document term matrix determined in the previous stage is divided into three different matrices at this stage. These are the left singular vector matrix (U), the singular value matrix (S), and the right singular matrix (V). It can be expressed by the following Eq. (1)'. The detailed description of Eq. (1) can be expressed as in Fig. 2.

$$A = USVT \tag{1}$$

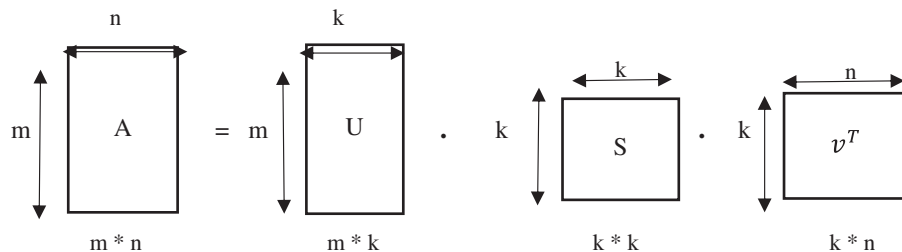


Figure 2: Singular value decomposition SVD

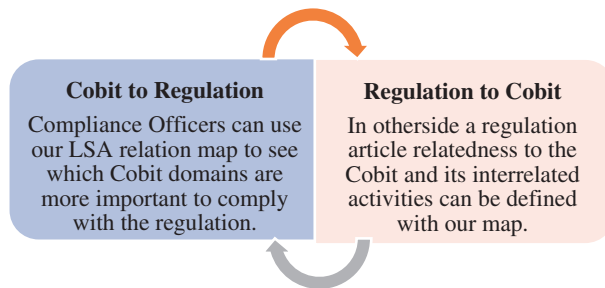


Figure 3: Map’s purpose; Cobit to regulation and regulation to Cobit

As shown in Fig. 2, it is obtained from the matrix U, which is an m * k-dimensional matrix and an n * k-dimensional V matrix, as shown in U and V. Thus, the situation in Eq. (2) can be reached [21].

$$UTU = VTV = 1 \tag{2}$$

As stated in the previous equations, there is a matrix representing each letter. The S matrix in Eq. (2) is expressed as $k \times k$. The matrix A can be expressed as the singular of the matrix S. At this point, for a better understanding of SVD, we can express matrix A as a difference matrix. If the column vectors of the U matrix are expressed as u_1, u_2, \dots, u_k , the diagonals of the S matrix as $\sigma_1, \sigma_2, \dots, \sigma_k$, and the column vectors of the V matrix as v_1, v_2, \dots, v_k , the resulting matrix A can be expressed as follows.

$$A = \sum_{i=1}^k \sigma_i u_i v_i^T \quad (3)$$

As expressed in Eq. (3), we can best use these equations for matrix A. For matrices σ_i and v_i^T we construct matrices such that $i = 1, 2, \dots, k$. Thus, using SVD, a matrix can be written as the sum of its components.

After that, SVD can now define and edit dimensions that express which data variations are included and how often in the text. At this stage, SVD takes the term document matrix. We now have an SVD vector to use to calculate similarity [22].

3.1.4 Calculating Cosine Similarity

This similarity is used to calculate the value between the document vector and the term vector in a database [17,18]. It is desirable that this value be small because the smallness of this ratio indicates the quality of the similarity level. The cosine similarity Eq. (4) is shown below.

$$\text{Cos}\alpha = \frac{A \cdot B}{|A| \cdot |B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2 \times \sum_{i=1}^n (B_i)^2}} \quad (4)$$

As expressed in the equation, A is the document and B is the term vector. Vector B can be thought of as the input of vector A. |A| is the length of vector A, and |B| is the length of vector between |A| and B. Thus, to be more precise, |B|, |A| It is a cross product between |B| and α can be expressed as the bridge between vector A and vector B [21].

3.2 Fuzzy AHP Expressions

One of the most studied methods of multi-criteria decision-making approaches (AHP) is. The analytical hierarchy process is a general theory of measurement. AHP is the broadest application used in multi-criteria decision making, planning and resource allocation, and conflict resolution [18]. AHP, put forward by Saaty, is a method that facilitates the hierarchical solution of multi-qualified, multilateral and multi-period structural problems. Although its purpose is to evaluate expert knowledge, the classical AHP method cannot respond to the ambiguity of human thinking. Therefore, fuzzy AHP is recommended for the solution of such problems [23].

The LSA method, which is the subject of our study, reveals the relationship between Cobit 2019 items, which express control targets in information technologies, and BRSA legislation, which expresses the supervision and regulation statements of the Turkish banking sector, and creates a similarity map. We aim to express the resulting analysis results in a more powerful way. Therefore, we used fuzzy AHP method, which is frequently used in recent studies, allows selection by experts for the criteria expressed in Tab. 1. Our 5 experts are having a deep understanding of both selected regulation and Cobit 2019. Compliance officers of a Turkish bank's whole universe were used in this study. They all have more than 10+ years of banking sector experience and have Cobit and ISACA

certifications. The determining criteria are the subject of our study, including the similarity between Cobit 2019 and the relevant regulation, and the basic domain detail headings of Cobit 2019 by ISACA.

Table 1: Criteria

C1	Area similarity
C2	Main domain similarity
C3	Subdomain similarity
C4	Activity/Article subject similarity
C5	Objective similarity

The reason why we used the fuzzy version of the AHP method in our study is that this method expresses the opinions of expert decision-makers more objectively than the AHP method. The fuzzy triangular numbers were used in this version used, which allows decision-makers to express their expressions more accurately. These numbers are indicated in [Tab. 2 \[23\]](#).

Table 2: Linguistic scales for fuzzy AHP

Rank	Linguistic term abbreviation	Triangular fuzzy number
1	Absolutely low importance	(1, 1, 2)
2	Essentially low importance	(1, 2, 3)
3	Weakly low importance	(2, 3, 4)
4	Equally low importance	(3, 4, 5)
5	Exactly equal	(4, 5, 6)
6	Equally high importance	(5, 6, 7)
7	Weakly high importance	(6, 7, 8)
8	Essentially high importance	(7, 8, 9)
9	Absolutely high importance	(8, 9, 9)

In order to apply the FAHP method, the opinions of 5 decision-makers who are well aware of both Cobit 2019 and the regulation were taken. Our sample was chosen from a Turkish Bank and they are the main universe of a compliance office. The application steps of this method are expressed as follows. Due to constraints, not all equations are expressed.

Step 1–2: In this step, pairwise comparison matrices expressed with fuzzy triangular numbers are created by 5 experts, and these matrices are turned into a single matrix with the geometric mean.

Step 3–4: In this step, l_i , m_i , u_i values will be obtained and normalization will be done in order to calculate the fuzzy weights. Then, the intersection points of l_i , m_i , u_i values and triangular fuzzy numbers will be determined by making pairwise comparisons [24].

$$V(M_2 \geq M_1) = \sup_{y \geq x} [\min(\mu_{M_1}(x), \mu_{M_2}(y))] \quad (5)$$

$$\mu_{M_2}(d) = \left\{ \begin{array}{ll} 1 & , m_2 \geq m_1 \\ 0 & , l_1 \geq l_2 \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} & , \text{in other cases} \end{array} \right\} \quad (6)$$

Step 5–6: In these steps, weight vectors will emerge, which are calculated as a result of pairwise comparisons. In the last step, the weight vectors will be normalized and sorted according to the criterion with the highest weight.

$$W' = (d'(A_1), d'(A_2), \dots, d'(A_n))^T \quad (7)$$

$$W = (d(A_1), d(A_2), \dots, d(A_n))^T \quad (8)$$

The resulting weight vector W is no longer a fuzzy number after normalization [25].

3.3 WASPAS Expressions

Using the criteria weight values obtained by the fuzzy AHP method, 25 randomly selected regulation statements from our matrix created by the one to many LSA method, with the highest correlation, 15 Cobit 2019 activities, and 10 lowest related Cobit 2019 activities were matched. Bilateral evaluation documents were created. Experts, who made the criteria determination in FAHP, also made WASPAS (Weighted Aggregated Sum Product Assessment) surveys voted for these 25 regulation pairs according to the score scale between 1 to 9. Our aim here is to determine whether the highest and lowest similarity rates determined by LSA are parallel with the result of the WASPAS method. Thus, it will be possible to compare the results of the LSA method with another method.

The WASPAS method was proposed by Zavadskas et al. [26]. This method has been developed as a combination of WSM (Weighted Sum Model) Weighted Sum Model and WPM (Weighted Product Model) Weighted Product Models. As a new method, WASPAS has been proposed as the most appropriate MCDM method for accuracy-based or confirmation by using both methods together. The purpose of this method is to increase sequencing accuracy [27]. The application steps of the WASPAS method are indicated by the following equations. Due to space constraints, the steps are summarized and expressed.

Step 1–2: is normalized according to the benefit and cost criteria for the resulting decision matrix. Utility-structured criteria are criteria whose values are desired to be maximized by the decision-maker. Cost-structured criteria are the criteria whose values are desired to be minimized. Since all criteria are tried to be maximized in this study, normalization will be done over the benefit criterion. It is specified in Eq. (9).

$$\bar{x}_{ij} = \frac{x_{ij}}{\max_i x_{ij}} \quad (9)$$

Step 3–4–5: The normalized decision matrix is multiplied by the criterion weights obtained by the fuzzy AHP method, and the row sum is made and $Q_i^{(1)}$ is obtained. It is specified in Eq. (10). Then, over the normalized decision matrix, each i . For the value of the alternative criterion, the power of the relevant criterion weight is taken and the values found are multiplied for each alternative, respectively, and the $Q_i^{(2)}$ value is calculated. It is specified in the Eq. (11). Finally, the combined optimality value is calculated for each alternative. This value obtained using Eq. (12) is calculated by considering the

results of the Weighted Sum Model and the Weighted Product Model [28].

$$Q_i^{(1)} = \sum_{j=1}^n \bar{x}_{ij} W_j \quad (10)$$

$$Q_i^{(2)} = \prod_{j=1}^n (\bar{x}_{ij})^{W_j} \quad (11)$$

$$Q_i = \lambda Q_i^{(1)} + (1 - \lambda) Q_i^{(2)} \quad (12)$$

Here; λ is the combined optimality coefficient and $\lambda \in (0, 1)$. In cases where the Weighted Sum Model and Weighted Product Model approaches have equal effects on the combined optimality criterion, $\lambda=0.5$ is taken. Thus, each alternative is ranked considering its combined optimality value Q_i . The alternative with the largest Q_i value is the best alternative and takes first place.

4 Applications

4.1 LSA Figures and Tables

The starting point of the LSA is text collections, within the text material usually, paragraphs are split up and create documents. The paragraphs saved in the documents have information about word relationships and can be represented abstractly in a frequency matrix, where the columns contain the individual documents and the lines that contain different words. In our research, regulation's every action and Cobit practices create documents and terms. We use one to many approaches stated on Colorado University LSA application which defined in Landauer and Dumains' paper [29].

The frequency of occurrence of a word in a specific document specifies the relatedness, this uses large corpora of natural language, then this frequency matrix is very sparse. The frequency matrix already contains all information about word relationships. In documents and texts there are usually too big and they are able to carry out unnecessary information ("noise"). In order to eliminate the "background noise", reducing the information contained in the frequency matrix to the core content. An essential step needed to create a filter of potentially superfluous words, application of weighting functions of the cell frequencies, singular value decomposition, and determination of the optimal number dimensions [30].

In the first step, we define the potentially unnecessary words and they are excluded. These include high-frequency words that do not convey any specific information, as well as words that appear very rarely, for example, less than three times in the entire text corpus. This reduces the number of different one's words clearly and increases the quality of the documents.

In our documents we have lots of IT-related special abbreviations, we completely changed these words with the help of a word processor. For example, we changed all IT abbreviations words to information technology. Cobit has lots of prepositions on it like, "e.g.", "ex.", "aka." or abbreviations like "I&T", "decision-making", "IT-enabled or (IT-something else)", "enterprise", "stakeholder", apostrophe or hyphen words are converted. Cobit gives too many examples on bracelets, these are deleted. Also, sometimes the example is off topic may be related to the document but it can affect the main area or direction and relatedness. So, we cleared these examples from actions and Cobit practices. This data evaluation increases the similarity ratio if any. On the other side, our sample regulation has some special country-specific information and some unrelated statements like revision purpose, the basis of legislation, definitions, abbreviations, and final provisions.

Cobit has 5 main domains and 40 processes (subdomain), these 40 processes have 231 practices and every practice has many actions as a piece of advice to comply with Cobit. Totally Cobit has 1202 actions. We got these 1202 actions and BRSA’s “Information Systems and Electronics of Banks Regulation on Banking Services” regulation which has 224 articles and made a one-to-many latent semantic analysis to gather a relatedness map. With this map, we intended to have two ways to evaluate, depicted in Fig. 3. First is finding Cobit’s related article on regulation, second from the regulation side which Cobit domains are covered and which Cobit actions can be done to comply with the regulation.

While applying the LSA, a similarity map was created by choosing Cobit 2019 as the main text and the relevant regulation items as the comparison text. Then, the opposite was applied and merging was done as much as the number of database matrices.

After coding our regulation and Cobit 2019 actions, we started to analyze every article with 1202 action from Cobit and created a matrix 224*1202. LSA gave us every article’s relatedness ratio with Cobit’s selected domains’ practices’ action’s similarity percentage. Values are varied from 0 to 1, 0 shows relatedness is not found and 1 shows they are the same. After analyzing all articles’ relatedness, we store data in Power BI table and create a heat map depicted in Fig. 3. to show easily which action is mostly related to which articles. In Fig. 4 each “Text” represents regulation statements and columns numbered with Cobit domain name and direct unique number of related activities.

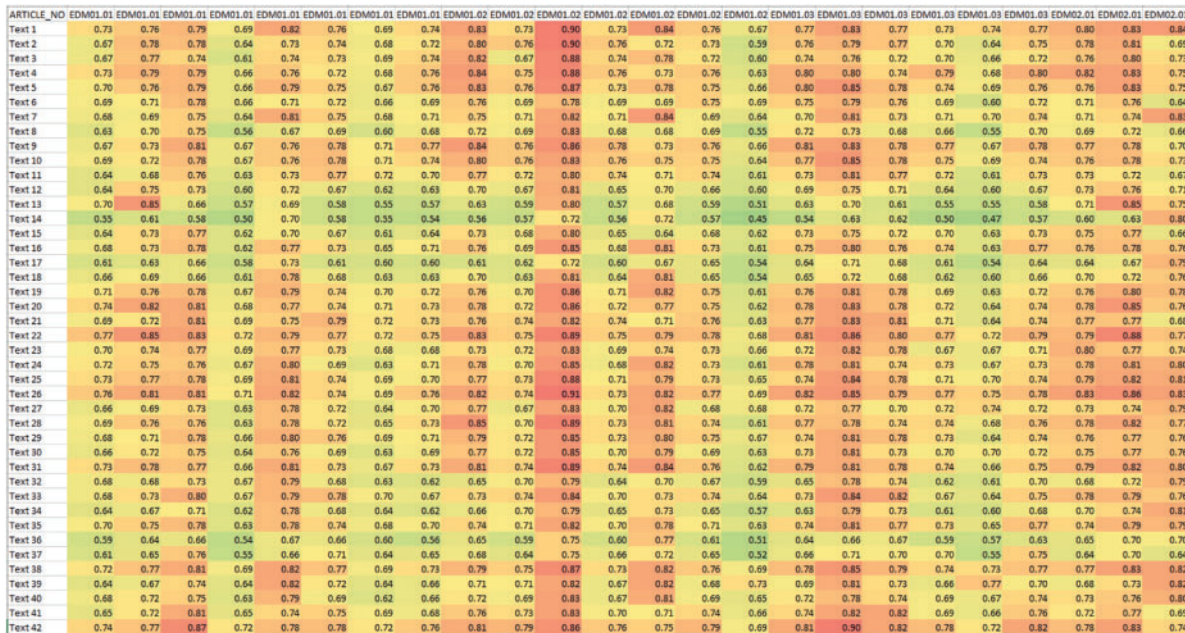


Figure 4: Regulation articles and Cobit actions relatedness map sample

To show our LSA results meaningfully we create a Power BI application with power query. Our dataset comes from the analysis result of our regulation articles and Cobit actions relation map. Data modal is shown on Fig. 5.

4.2 Fuzzy AHP Tables

The criterion weights obtained after the applied fuzzy AHP steps are expressed in Tab. 3.

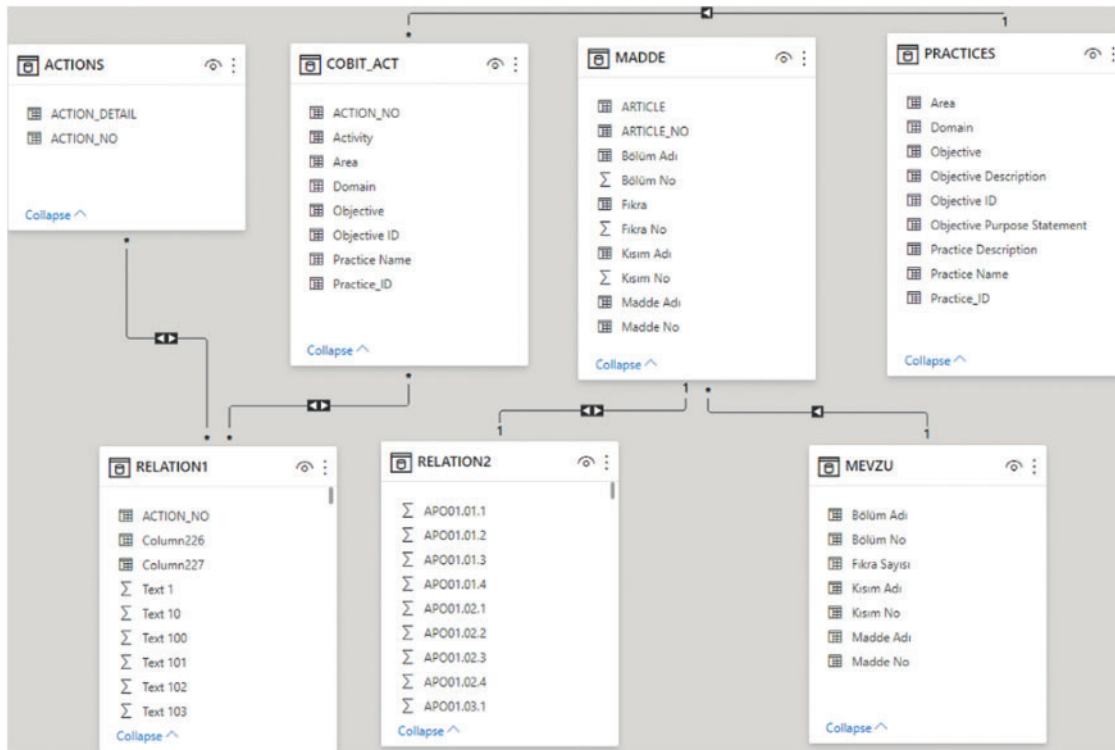


Figure 5: Power BI data modal to query by experts

Table 3: Criteria weights

Symbol	Criteria	Weight	Rank
C1	Area similarity	0.0086	5
C2	Main domain similarity	0.0296	4
C3	Subdomain similarity	0.0677	3
C4	Activity/Article subject similarity	0.3060	2
C5	Objective similarity	0.5882	1

After the LSA method was applied, the fuzzy AHP method was used to strengthen our study. According to the results of this method, among the determining criteria, Practice/Article Subject Similarity (C4) and Objective Similarity (C5) were determined as the criteria with the highest criterion weights. This means that, according to the opinions of 5 experts who dominate all regulations, Cobit 2019 and related regulation items have 58 percent similarity in purpose, 30 percent application and subject similarity. These results also support our LSA method.

4.3 WASPAS Tables

Ranking results are expressed in Tab. 4.

According to the results in Tab. 4, all comparison pairs Q1 (+)–Q14 (+), which were determined to have the highest similarity in the LSA method, were also ranked as the highest similarity according

to the WASPAS method. Likewise, Q15 (-)–Q25 (-) with the lowest similarity ratio according to the LSA method was determined as the comparison pairs with the lowest similarity ratios according to the WASPAS method.

Table 4: Ranking regulation and Cobit activity pairs according to WASPAS and LSA results

Symbol	WASPAS value	Rank	LSA results
Q12 (+)	0.97	1	0,92
Q4 (+)	0.93	2	0,91
Q2 (+)	0.93	3	0,90
Q10 (+)	0.93	4	0,89
Q7 (+)	0.92	5	0,86
Q1 (+)	0.92	6	0,90
Q13 (+)	0.91	7	0,87
Q8 (+)	0.91	8	0,85
Q5 (+)	0.90	9	0,86
Q6 (+)	0.90	10	0,89
Q11(+)	0.90	11	0,92
Q3 (+)	0.89	12	0,90
Q14 (+)	0.87	13	0,91
Q9 (+)	0.86	14	0,90
Q16 (-)	0.28	15	0,26
Q18 (-)	0.27	16	0,31
Q17 (-)	0.26	17	0,32
Q19 (-)	0.25	18	0,29
Q24 (-)	0.24	19	0,27
Q21 (-)	0.23	20	0,30
Q15 (-)	0.23	21	0,25
Q22 (-)	0.22	22	0,33
Q20 (-)	0.19	23	0,32
Q25 (-)	0.19	24	0,28
Q23 (-)	0.19	25	0,29

5 Conclusion

Nowadays businesses are operating in a highly regulated areas and their thread landscape is generally high. To govern information technologies, they should create mechanisms to have effective oversight of IT operations. To create effective governance, businesses use best practices and regulations. Best practice compliance is not important as regulations. To become more mature in both areas the requirement is much to be compliant with regulations and some selected best practices. Creating a regulation relevance map with a best practice creates value to practitioners. Compliance officers and

also governing body always trying to find the most relevant path between related best practices to easily understand and generalize or exemplify the regulation statements. This complexity may create wrong interpretations and commentary about specific issues in regulation. Also, another aspect about best practice mapping is creating only master domain relatedness by best practice owner, this is not enough for governing bodies. They need to understand not only master domains but also any area, subdomains, objectives, and practices.

Our LSA-based relatedness method facilitates the regulation to best practices or best practice to regulation relation. Searching the whole similarity between all practices and every statement is creating big data for analyze. This dataset can be used for finding any selected regulation statement and specific related best practice practices. The most challenging step is cleaning each of the documents to get valid results. LSA's one to many approaches gets every regulation statement one by one and every practice (in our sample 1202 Cobit statement) relation with cosine similarity method. With the help of LSA method, we create every regulation relatedness percentage to any of Cobit practices. Practitioners can also use these matrices' transpose to create Cobit practice relatedness to any regulation statement. In our sample we create a Power BI report to easily apply this two-way analysis.

Finally, we used FAHP and WASPAS methods to verify our relatedness matrices. Our results show that LSA method creates valid correct similarities and can be usable in regulation and best practice mapping and creating relatedness matrices. Creating a relatedness map between a regulation and a best practice in an objective way is possible and creates consistency by the help of LSA. Sample bank can easily adapt its processes to find Cobit to regulation related articles or regulations' related Cobit activities easily. Moreover, not only in our sample bank but also in all sector representatives can use this modal, due to they are all subject to same regulations and best practices.

6 Recommendations

After creating an LSA-based relatedness map, a business should create a project or a program that addresses the need of increasing the maturity level of the compliance process. This project may focus on techniques to assess results with the related regulation and best practices by the sponsorship of governing body. In this way, regulation findings, issues, and risks can be addressed and can be solved in a collaborative environment.

Our research covers the relatedness of two standards; Cobit and BRSA's IT-related regulation. Businesses also want to implement other standards like ISO 27001 or NIST (National Institute of Standards and Technology) standards and there are too many regulations to be compliant. In the future, researchers can use our LSA-based method to create a relatedness map with each standard or regulation pair. Also, we can implement ready-to-use relatedness maps to bridge our relatedness map.

Funding Statement: The author received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] R. Sheikhpour and N. Modiri, "An approach to map Cobit processes to ISO/IEC 27001 information security management controls," *International Journal of Security and Its Applications*, vol. 6, no. 2, pp. 13–28, 2012.

- [2] W. Gomaa and A. Fahmy, "A survey of text similarity approaches," *International Journal of Computer Applications*, vol. 68, no. 13, pp. 13–18, 2013.
- [3] M. Lintean, C. Moldovan, V. Rus and D. McNamara, "The role of local and global weighting in assessing the semantic similarity of texts using latent semantic analysis," in *23rd Int. FLAIRS Conf.*, Florida, USA, pp. 235–240, 2010.
- [4] B. Stone, S. Dennis and P. J. Kwantes, "Comparing methods for single paragraph similarity analysis," *Topics in Cognitive Science*, vol. 3, no. 1, pp. 92–122, 2011.
- [5] A. Dai, C. Olah and Q. Le, "Document embedding with paragraph vectors," *Cornell University Preprint Arxiv*, vol. 1507.07998, pp. 1–8, 2015.
- [6] R. Almeida, R. Lourinho, M. Silva and R. Pereira, "A model for assessing Cobit 5 and ISO 27001 simultaneously," in *IEEE 20th Conf. on Business Informatics (CBI)*, Vienna, Austria, vol. 1, pp. 60–69, 2018.
- [7] F. Fathoni, N. Simbolon and D. Hardiyanti, "Security audit on loan debit network corporation system using Cobit 5 and ISO 27001: 2013," in *Journal of Physics: Conference Series*, vol. 1196, no. 1, Palembang, Indonesia, pp. 1–8, 2019.
- [8] S. Zhang and H. Fever, "An examination of the practicability of Cobit framework and the proposal of a Cobit-BSC model," *Journal of Economics, Business and Management*, vol. 1, no. 4, pp. 391–395, 2013.
- [9] R. Almeida, P. L. Pinto and M. M. Silva, "Using archimate to assess cobit 5 and ITIL implementations," in *25th Int. Conf. on Information Systems Development*, Katowice, Poland, pp. 235–246, 2016.
- [10] L. Elluri, K. P. Joshi and A. Kotal, "Measuring semantic similarity across EU GDPR regulation and cloud privacy policies," in *IEEE Int. Conf. on Big Data (Big Data)*, Atlanta, USA, pp. 3969–3978, 2020.
- [11] S. Hassan and R. Mihalcea, "Semantic relatedness using salient semantic analysis," in *55th AAAI Conf. on Artificial Intelligence*, San Francisco, USA, pp. 884–889, 2011.
- [12] M. Louwerse, Z. Cai, X. Hu, M. Ventura and P. Jeuniaux, "Cognitively inspired NLP-based knowledge representations: Further explorations of latent semantic analysis," *International Journal on Artificial Intelligence Tools*, vol. 15, no. 6, pp. 1021–1039, 2006.
- [13] M. Baroni, G. Dinu and G. Kruszewski, "Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors," in *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, USA, 2014.
- [14] P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics," *Journal of Artificial Intelligence Research*, vol. 37, no. 1, pp. 141–188, 2010.
- [15] E. Gabrilovich and S. Markovitch, "Computing semantic relatedness using Wikipedia-based explicit semantic analysis," in *Int. Joint Conferences on Artificial Intelligence*, San Francisco, CA, USA, vol. 7, pp. 1606–1611, 2007.
- [16] T. K. Landauer, D. Laham and R. Rehder, "How well can passage meaning be derived without using word order? A comparison of latent semantic analysis and humans," in *Proc. of the 19th Annual Meeting of the Cognitive Science Society*, Stanford, CA, USA, pp. 412–417, 1997.
- [17] W. Pan, X. Xu and C. K. Chang, "Clustering mashups by integrating structural and semantic similarities using fuzzy AHP," *International Journal of Web Services Research*, vol. 18, no. 1, pp. 34–57, 2021.
- [18] G. Cosma and M. Joy, "An approach to source-code plagiarism detection and investigation using latent semantic analysis," *IEEE Transactions on Computers*, vol. 61, no. 3, pp. 379–394, 2012.
- [19] S. Deerwester, S. T. Dumais and R. Harshman, "Indexing by latent semantic analysis." *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990.
- [20] A. Luthfiarta, J. Zeniarja and A. Salam, "Latent semantic analysis (LSA) algorithm for automatic document summarization for document clustering." in *the National Seminar on Applied Information and Communication Technology*, Semarang, Indonesia, pp. 13–18, 2013.
- [21] W. Darmalaksana, C. Slamet, W. Zulfikar, I. F. Fadillah, D. S. Maylawati *et al.*, "Latent semantic analysis and cosine similarity for Hadith search engine," *Telkommika Telecommunication, Computing, Electronics and Control*, vol. 18, no. 1, pp. 217–227, 2020.

- [22] T. L. Saaty, "Decision making with the analytic hierarchy process," *International Journal of Services Sciences*, vol. 1, no. 1, pp. 83–98, 2008.
- [23] G. Büyüközkan and G. Çifçi, "A combined fuzzy AHP and fuzzy TOPSIS based strategic analysis of electronic service quality in healthcare industry," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2341–2354, 2012.
- [24] D. S. Maylawati, W. B. Zulfikar, C. Slamet, M. A. Ramdhani and Y. A. Gerhana, "An improved of stemming algorithm for mining Indonesian text with slang on social media," in *6th Int. Conf. on Cyber and IT Service Management*, IEEE, pp. 1–6, 2018.
- [25] M. Madić, V. Gecevska, M. Radovanović and D. Petković, "Multi-criteria economic analysis of machining processes using the WASPAS method," *Journal of Production Engineering*, vol. 17, no. 2, pp. 1–6, 2014.
- [26] E. K. Zavadskas, J. Antucheviciene, J. Šaparauskas and Z. Turskis, "Multi-criteria assessment of facades' alternatives: peculiarities of ranking methodology," in *11th Int. Conf. on Modern Building Materials, Structures and Techniques*, Vilnius, Lithuania, vol. 57, no. 112, pp. 107–112, 2013.
- [27] J. Šaparauskas, E. K. Zavadskas, and Z. Turskis, "Selection of facade's alternatives of commercial and public buildings based on multiple criteria," *International Journal of Strategic Property Management*, vol. 15, no. 2, pp. 189–203, 2011.
- [28] Z. Turskis, E. K. Zavadskas, J. Antucheviciene and N. Kosareva, "A hybrid model based on fuzzy AHP and fuzzy WASPAS for construction site selection," *International Journal of Computers Communications & Control*, vol. 10, no. 6, pp. 113–128, 2015.
- [29] T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge," *Psychological Review*, vol. 104, no. 2, pp. 211–240, 1997.
- [30] W. Lenhard, H. Baier, J. Hoffmann and W. Schneider, "Automatische bewertung offener antworten mittels latenter semantischer analyse," *Diagnostica*, vol. 53, no. 3, pp. 155–165, 2007.