Preprint: In proceedings of IEEE International Services Computing Conference (SCC) 2021

Analyzing GDPR compliance in Cloud Services' privacy policies using Textual Fuzzy Interpretive Structural Modeling (TFISM)

Ronak Razavisousan
Information Systems Department
University of Maryland Baltimore County (UMBC)
Baltimore, United States
Ronak2@umbc.edu

Karuna P. Joshi
Information Systems Department
University of Maryland Baltimore County (UMBC)
Baltimore, United States
karuna.joshi@umbc.edu

***Abstract*— Cloud Service providers must comply with data protection regulations, like European Union (EU) General Data Protection Regulation (GDPR), to ensure their users' personal data security and privacy. Hence, the service privacy policies and terms of service documents refer to the rules it complies with within the data protection regulation. However, these documents contain legalese jargon that requires significant manual effort to parse and confirm compliance. We have developed a novel methodology, Textual Fuzzy Interpretive Structural Modeling (TFISM), that automatically analyzes large textual datasets to identify driving and dependent factors in the dataset. TFISM enhances Interpretive Structural Modeling (ISM) to analyze textual data and integrate it with Artificial Intelligence and Text extraction techniques. Using TFISM, we identified the critical factors in GDPR and compared them with various Cloud Service privacy policies. In this paper, we present the results of this study that identified how different factors are emphasized in GDPR and 224 publicly available service privacy policies. TFISM can be used both by service providers and consumers to automatically analyze how close a service privacy policy aligns with the GDPR.**

Keywords*— *GDPR, Privacy Policy, User's privacy, Fuzzy Interpretive Structural modeling (FISM), Textual Fuzzy Interpretive Structural Modeling (TFISM)

I. INTRODUCTION

The extensive penetration of Cloud-based services has increased concerns about personal privacy and data sharing. Consumer data collected by service providers can include browsing history, shopping history, geographic location, Personally Identifiable Information (PII) (such as name, address, phone number, etc.), and other related data. Data protection laws are being implemented by authorities and standards bodies globally to address consumer apprehensions. One key data protection regulation is General Data Protection Regulation (GDPR) [1] that focuses on data protection and privacy in the European Union's (EU) and the European Economic Area (EEA). It also addresses the transfer of personal data outside the EU and EEA. Although the various laws and regulations are not comprehensive and unified, they all have strict requirements about deleting the online data collected from consumers.

All service providers must adhere to data protection regulations and specify them in their service's privacy policies. These privacy policies often shared publicly on the provider's website describe how the provider collects, uses, and manages or shares data. The privacy policy also makes it easier for service providers to collaborate with other companies or third parties. However, these policies are only available as large text documents that are not machine-interpretable. Moreover, they contain legalese jargon that is difficult to comprehend. A significant manual effort by a legal expert is currently required to determine how closely a service privacy policy adheres to a specific data protection regulation.

We have developed a novel methodology, called Textual Fuzzy Interpretive Structural Modeling (TFISM), that identifies vital terms or influential factors in a textual dataset, prioritizes the power of each factor, and then determines the connections and hierarchies between these factors. TFISM is a domain-free methodology that can assist in complex decision-making. We applied this methodology to GDPR and some publicly available privacy policies of 224 Cloud-based Service vendors. We were able to compare how closely the two documents prioritized different key terms or factors. This comparison gives us an insight into how the same concepts are emphasized differently in an authority document and a referring document. In this paper, we present our TFISM methodology and the results of our study with the service privacy documents.

The rest of this paper has been organized as follows: Section II covers the Related Work. Section III describes the TFISM methodology. Section IV describes the results of our study. We conclude in Section V.

II. RELATED WORK

A. Interpretive Structural Modeling (ISM)

ISM is a methodology to identify the relationship between variables in the decision-making process or complex problem-solving. ISM was invented by Warfield [2] for "interactive management" and "structural approach" to system design. ISM has an enhanced version called Fuzzy-ISM [3] which was mentioned by Ragade for the first time. Fuzzy-ISM uses Fuzzy logic, which plays a significant role in dealing with vagueness and uncertainty in human language and thoughts in decision making. To integrate various opinions, experiences, ideas, and motivations of individual experts, it becomes essential to translate the linguistic judgments into fuzzy numbers [4]. The integration of ISM with fuzzy sets provides flexibility to decision-makers to further understand the level of influences of one criterion over another, which was earlier present only in the

form of binary (0,1) numbers [5]. We extended Fuzzy-ISM by enabling textual data as input.

B. MICMAC Analysis

The Matrice d'impacts croisés multiplication appliquée á un classment (MICMAC) method was developed by Michel Godet and François Bourse [6] and means Cross-Impact Matrix Multiplication Applied to Classification. MICMAC is used to examine the relationship strength between significant system factors based on their driving and dependency power. We can say it is a cross-impact matrix multiplication applied to classification to study indirect relationships [7]. MICMAC analysis complements the ISM method by visualizing the relationships between variables in a complete view of the system [8]. MICMAC analysis divides the plot into four sections. The autonomous section located at the left-Bottom shows the factors that are not closely related to the system and have weak or no dependence on other factors. Section two refers to dependent factors in the right-bottom that are the factors that are primarily dependent on other factors. Linkage factors are presented in the third section that refers to the connecting factors that are unstable and most influence others. Linkage factors are in the top-right section. The fourth section in the top-left is independent factors. These factors have weak influence from other factors and must be paid maximum attention owing to the strong key factors.

We have integrated MICMAC analysis into our methodology as described in section III.

C. Study of GDPR and Service Privacy Policy

Privacy policies are critical for helping users make informed decisions about their data and sharing it. Therefore, many studies are focused on identifying the challenges in understanding a privacy policy. Linden et al. [9] leveraged machine learning techniques to classify more than 6000 privacy policies. They were able to highlight the changes in the privacy policies before and after the GDPR release. Zaeem et al [10] employed Natural Language Processing (NLP) techniques to automatically analyze the privacy policy content from three corpora. They manually labeled the data with pre-and-post-GDPR and compared the results with GDPR to identify the role of GDPR.

For addressing users' concerns about privacy policies based on different requirements, Chang et al. proposed the automated privacy policy extraction system, which has the function to customized based on user's concerns. [11] then they compare their results from a transparency point of view with GDPR. Tesfay et.al [12] proposed a machine learning-based approach to address the issue of understanding the content and details of the privacy policy by users. They summarize the long privacy policy into short and condensed notes following a risk-based approach. They used GDPR to validate their approach. GDPR used as a standard model to compare in research has also been done by Torre, et al. [13]. They proposed a conceptual model to characterize the content of privacy policies and then classify them to compare with GDPR easily. This approach informs the users about the summary of the content.

Our approach to this problem of comparing context in the two documents referencing each other has been more multi-disciplinary since we have integrated concepts from Management Science, like Fuzzy ISM and MICMAC analysis, with techniques from Artificial Intelligence and Information Retrieval.

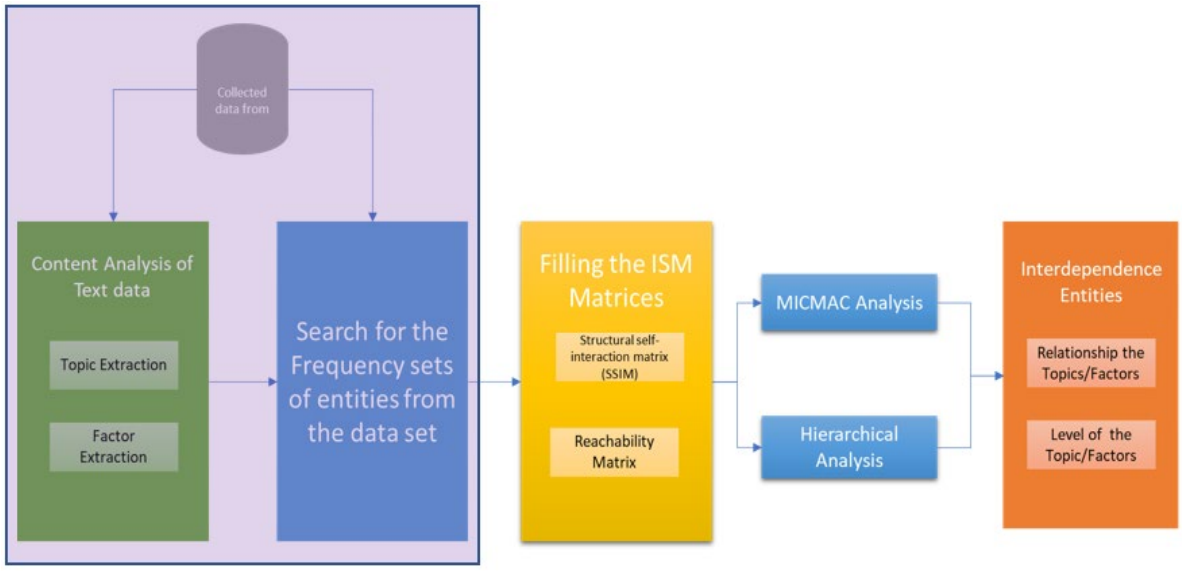

FIGURE 1: HIGH LEVEL ARCHITECTURE OF THE TEXTUAL FUZZY INTERPRETIVE STRUCTURAL MODELING

III. METHODOLOGY

We have developed the Textual Fuzzy Interpretive Structural Modeling (TFISM) methodology to identify the influential factors in a textual dataset, prioritize the power of each factor, and determine the connections and hierarchies between these factors. Our method extends Interpretive Structural Modeling (ISM) by engaging text data as input data to build a decision-making model that is not dependent on any prior knowledge or human intuition. TFISM is a domain-free methodology and can facilitate complex decision-making.

Each problem or complex situation relates to some causing variables or resolving factors. These variables have different roles that can distress decision-making. Sometimes they have interrelationships that can cause changes in each other. Considering all effects and their relationships is critical for analyzing situations and making decisions. ISM [2] is an established methodology for identifying relationships among specific items, which define a problem or an issue. The limitation of the ISM method is its reliance on experts' knowledge in each area. To be more specific, some matrices are manually filled based on the expert's opinion, while in our proposed method the process is adjusted to substitute text data for human intuition. TFISM includes all the features in the ISM, moreover enabling us to take advantage of the huge amount of textual documents.

Create Structural Self-Interaction Matrix (SSIM) is the first step of running the ISM method. The SSIM matrix is developed for variables, which indicates pair-wise relationships among variables of the system under consideration. Figure 2 The right-hand shows an example of SSIM that is manually field. Different types of relationships are presents as A, O, V, and X. On the left side, the same matrix has been presented, which is filled by the TFISM method. Each number demonstrates the frequency of appearance of a set of factors in the input resources. The process of interpreting the number to the relationships and transforming them to the visual data to understand is the novelty of the TFISM method which is presented as a different paper [14] in this research chain.

GDPR SSIM	data	subject	process	person	authority	right	protection	law	public	information
data		A	X	V	A	O	V	X	A	V
subject			V	O	A	A	A	A	V	X
process				V	X	X	V	O	V	A
person					O	X	V	A	V	V
authority						V	V	V	V	A
right							V	V	A	O
protection								X	O	A
law									V	A
public										O
information										

GDPR Text	data	subject	process	person	authority	right	protection	law	public	information
data	589	264	371	301	80	141	134	75	65	66
subject	466	293	273	214	79	131	68	65	62	61
process	434	185	480	243	90	106	98	67	62	52
person	466	178	298	313	49	108	91	64	47	48
authority	175	93	166	89	246	56	40	33	44	29
right	307	165	192	149	43	158	54	48	44	36
protection	256	86	164	119	37	58	141	38	39	21
law	194	108	156	112	50	60	36	113	46	29
public	174	91	121	105	39	51	45	46	100	24
information	201	91	103	117	25	38	29	21	21	97

Figure 2: Sample of SSIM filed by ISM and TFISM

Figure 1 shows the high-level architecture of the TFISM methodology. The highlighted part in the purple square shows the main contributions for implementing the ISM method for text data. TFISM is a method that transforms the ISM method by integrating a new process flow for using textual data as input [14]. This approach builds on the frequency of two words or two variables in a textual data set. As figure 1 illustrates TFISM method has four steps that are described below.

1. **Factor Determination:** Output of this step is a list of factors or variables in a dataset whose strength and the role we want to determine. Sometimes these factors are clearly defined, and TFISM will directly apply them to the next step. In a situation where the influential factors are not easily identifiable, techniques like topic extraction with the help of domain experts can be used to identify the factors or variables. Then for every two sets of variables, the frequency of adjacency will be determined. The adjacency can be identified in a sentence, paragraph, or whole document; one of each has been chosen based on the document length and its structures.

2. **Populating the Matrices**: In step two, the ISM matrices are populated. The approach to customizing this step for numeric variables with an unlimited input range is presented in the previous paper [14]. In this step, the frequency of every two factors will be normalized, and the output will calculate the effect of all factors on a specific factor.

3. **Performing MICMAC analysis and Hierarchical analysis**: Step three will compete by generating MICMAC analysis and Hierarchical analysis. As explained above, MICMAC analyses the driving power and dependency between two factors. In the TFISM approach, the summation of normalized power from the reachability matrix provides the strength of each factor. The column summation shows dependency power, and the row summation shows driving power. The combination of these values gives us a point in the MICMAC plot. The location of each factor in the plot determines the power of the factor to drive other factors, change the system, and show how dependent it is on other factors. Hierarchical analysis requires reachability set, antecedent set, and intersection sets. Reachability set consists of variables to which it may help achieve or reach, while the antecedent set consists of variables that may help achieve or reach it. The process for running a hierarchical plot was presented in our related paper [14]. Hierarchical plot levels the factors based on the effectiveness to change the system. It shows how an adjustment in a variable flows to the others and affects the whole system. In other words, it shows us for changing a system which variable should adjust and how the others will react. The hierarchical analysis is a part of the ISM method; we enhanced the process in ISM to use real numbers (R) instead of natural numbers (N).

4. **Interpreting the results:** Step four focuses on analyzing the results of step 3 to determine the levels of each factor and the relationships between the various factors. Details are explained in the results section IV.

A. Variable selection

The first step of our methodology is to determine the relevant factors for privacy policies. Relevant factors refer to the effective or influential factors that have power for changes from the perspective of our study. For example, data privacy for the medical domain has a critical role that cannot be ignored. On the other hand, for scientific datasets, like weather prediction, privacy may not be an influential factor. For the same domain, depending on the perspective of the study, the effective factors considered in TFISM may be different. If we do not choose an

appropriate list of factors, TFISM cannot determine the power of factors and their interrelationship. For selecting a comprehensive list of factors in privacy policies, several sources were analyzed.

Ontology of GDPR policy document: One of the useful techniques for having a 360-degree view of any concept is using its pre-existing Ontology or taxonomy. For this study, the GDPR ontology created by Elluri and Joshi [15] was analyzed. Since we are looking at the effective factors, the outlines of the high level of ontology were employed. The key factors in GDPR ontology identified by us are listed in Table 1. However, when we searched them in the privacy policy sources, we found out that these words are not frequent in the sources.

TABLE I. GDPR FACTORS SELECTED FROM THE ONTOLOGY

	Ontology Terms
1	Customer
2	Cloud provider
3	Cloud security controller
4	Incident management
5	Contingency plan
6	Computer access control
7	Datacenter service
8	Encryption
9	Cloud awareness and training
10	Identification and authentication
11	Cloud portability and interoperability
12	Cloud compliance audit

Hence, we extracted high-frequency terms and phrases in the sections and subsections of the GDPR document. In contrast to the ontology term list, the extracted list from GDPR was more detailed and covered common words found in the privacy policy text. However, the number of terms extracted was huge – we extracted 66 factors, making it difficult to do a TFISM analysis. The approach that helped us reduce the term list was topic modeling using LDA [16]. We applied the LDA to the whole GDPR document and extracted the frequent topics. Leveraging topic modeling had another advantage in that it shortened the phrases to a one-word topic. Our previous studies showed that looking for two words phrases did not provide reliable data for the TFISM analysis. So having factors as a single word was most suited for our methodology. With the help of an expert, we merged the topic list and GDPR outlines to ensure that the list of factors covered essential and critical factors of the privacy policy. At this process, we also shortened the long-phrases to one-word factors. The final list of 35 factors is shown in Table II.

We began the TFISM process with 35 factors of the final list and got some initial results. Still, the analysis was too crowded hence identifying the relationships between factors proved hard.
Although the results are readable and reliable, we decided to reduce the list by choosing 10 frequent factors to have precise results and plots to present. The final list of 10 frequent terms after filtering by the expert is listed in table III.

TABLE II: THE FINAL LIST OF GDPR FACTORS EXTRACTED BY AN EXPERT

No.	Factor	No.	Factor
1	Access	19	law
2	Accordance	20	Member
	Article	21	order
4	Audit	22	paragraph
5	Authorities	23	persons
6	Board	24	processor
7	Commission	25	protection
8	confidentiality	26	provider
9	controller	27	purposes
10	customer	28	rights
11	data	29	rule
12	decision	30	safeguards
13	encryption	31	security
14	freedoms	32	State
15	Health	33	storage
16	identification	34	subjects
17	information	35	Union
18	integrity		

TABLE III: FINAL LIST OF TERMS USED FOR ANALYSIS

No.	Factor	No.	Factor
1	subject	6	law
2	process	7	information
3	person	8	public
4	authority	9	protection
5	right	10	data

B. Privacy policy selection

Preparing the list of providers that offer cloud services was our next task. We considered a wide range of providers, including retail companies with online presence, entertainment companies with online services, companies for internet and web services, and governmental services. Having privacy policies from a wide range of companies enabled us to study and compare companies in the same domain. We were able to build a corpus of over 224 privacy policy documents from various service providers.

We considered the GDPR document to be the authoritative document that is referenced by the privacy policies. For our study, we included privacy policies from articles that are referred by studies and other companies [9], [10], [11], [12], and [13].

C. Data source comparison

The third step in this process is deciding how to use data sources to have the most reliable results. As mentioned, we have

two sets of documents. One is the GDPR document which is our reference and ground truth. The other set is the corpus of privacy policies which were our test data. We used the test data in two different ways.

1) Combination of all privacy policy documents

The large corpus of the privacy policies of the US companies encourages us to see how this concept is viewed in Europe compared to the US. For this aim, we merged all 224 policies into a single document and used it to compare with GDPR.

2) Individual privacy policies of service providers

We analyzed the privacy policies of each provider individually and compared them with GDPR to identify the factor differences and emphasizes. In this paper, we present the results of our study of the privacy policies of Amazon [16] and Walmart [18].

IV. Results

In this section, we present the results of our analysis. As mentioned, the analysis was done by comparing the factors in the GDPR document and a single combined document of all privacy

policies and the second analysis was of individual policies of the service providers Amazon [16] and Walmart [18]

A. Comparing the combination of all privacy policies with GDPR

MICMAC Analysis

Our first approach for comparing two datasets focused on the MICMAC analysis of two data sets. Figure 3 and figure 4 present the MICMAC plot based on GDPR data and the same result for the combination of all privacy policies. When we compare two MICMAC plots some interesting results were observed.

For the analysis, it is important to be aware of the values range if they are different. In this case, the GDPR has higher values than the other plot (All privacy policies), and it is indicated that selected factors have less priority in figure 4.

In the GDPR plot, there is a linkage variable which is "Person" while the combination of the privacy policy does not have any linkage variable. About the importance and power of linkage variable explained before, now it is interesting that a concept such as "person" is in this position.

Some variables like "Public", "Protection", and "data" have the same category in both plots. Some factors have opposite roles. For instance, "Right" in GDPR categorize as a driver factor while in all privacy policy is in the dependent factors. "Law "and "information" have the same situations which mean for GDPR they are driver and for privacy policies of companies are dependent. Previously we clarified the role of drivers and their power to push other factors to change. Look at the driver factors shows the emphasis of each data set. GDPR has "information", "Law", and "right" in the category of independent factors. Whereas "Authority" and "person" are in the same category for companies' privacy policies. It is noticeable that these recent factors have low driving power and they are very close to the autonomous section.

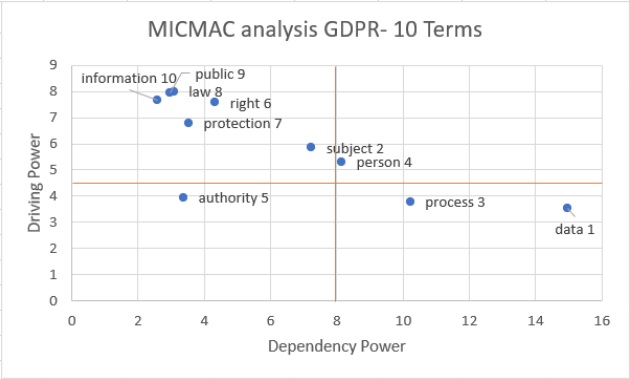

FIGURE 3: MICMAC PLOT - GDPR FOR 10 TERMS

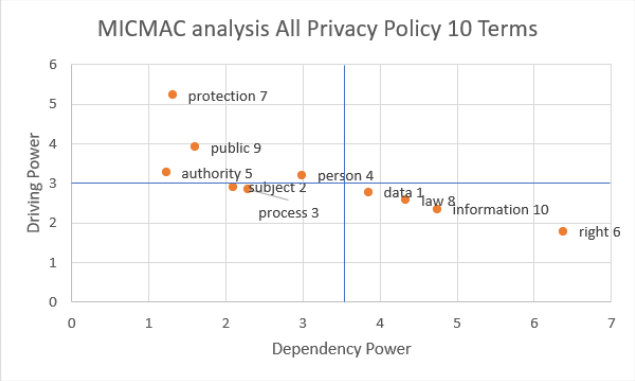

FIGURE 4: MICMAC PLOT - ALL PRIVACY POLICIES - 10 TERMS

We also propose another way for analyzing MICMAC data by comparing dependency power and driving power separately. Figure 5 shows the dependency power of two data sets and figure 6 presents the driving power of datasets. Generally, blue lines (which present the GDPR data) are above the orange lines because GDPR has a higher range of power. Interestingly in figure 5, we can see the points that the orange line comes over the blue line. These points refer to the "Authority", "Protection", and "Information". Although the combination of all privacy policies has less power these three factors are more dependent comparing to the GDPR. Although the orange line is more flattened in figure 6 it seems that the driving power in both data sets has a similar path. The "information" is an exception for driving power. While GDPR has the highest driving power for information, this factor in All privacy policies has very low driving power. There is a similar situation for the "Right" when comparing the driving power.

B. Comparing the Privacy policy of cloud services individually with GDPR

In this section, we present the results of the study when factors of a service privacy policy were separately analyzed and compared with factors in GDPR.

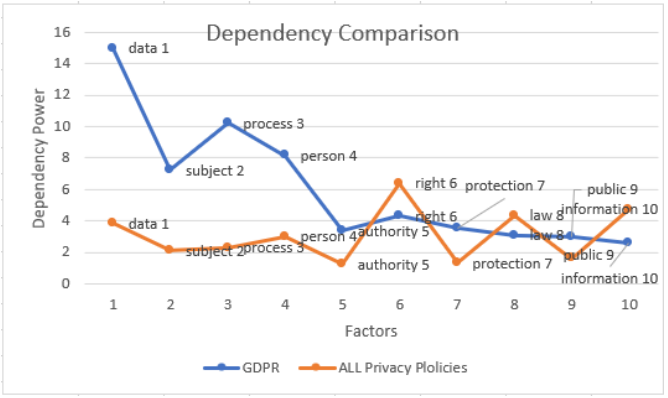

FIGURE 5: DEPENDENCY COMPARISON FOR TWO DATA SETS

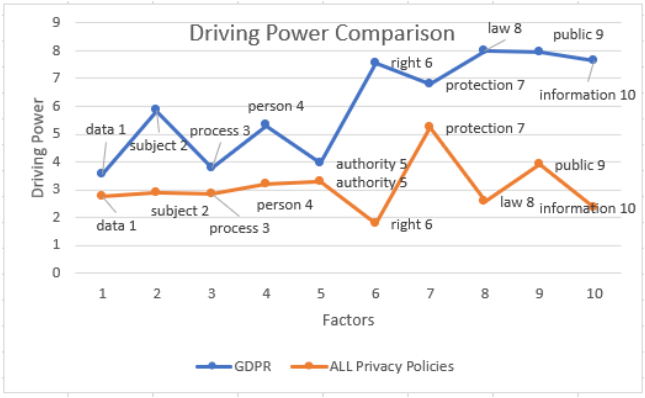

FIGURE 6: INDEPENDENCY (DRIVING POWER) COMPARISON FOR TWO DATASETS

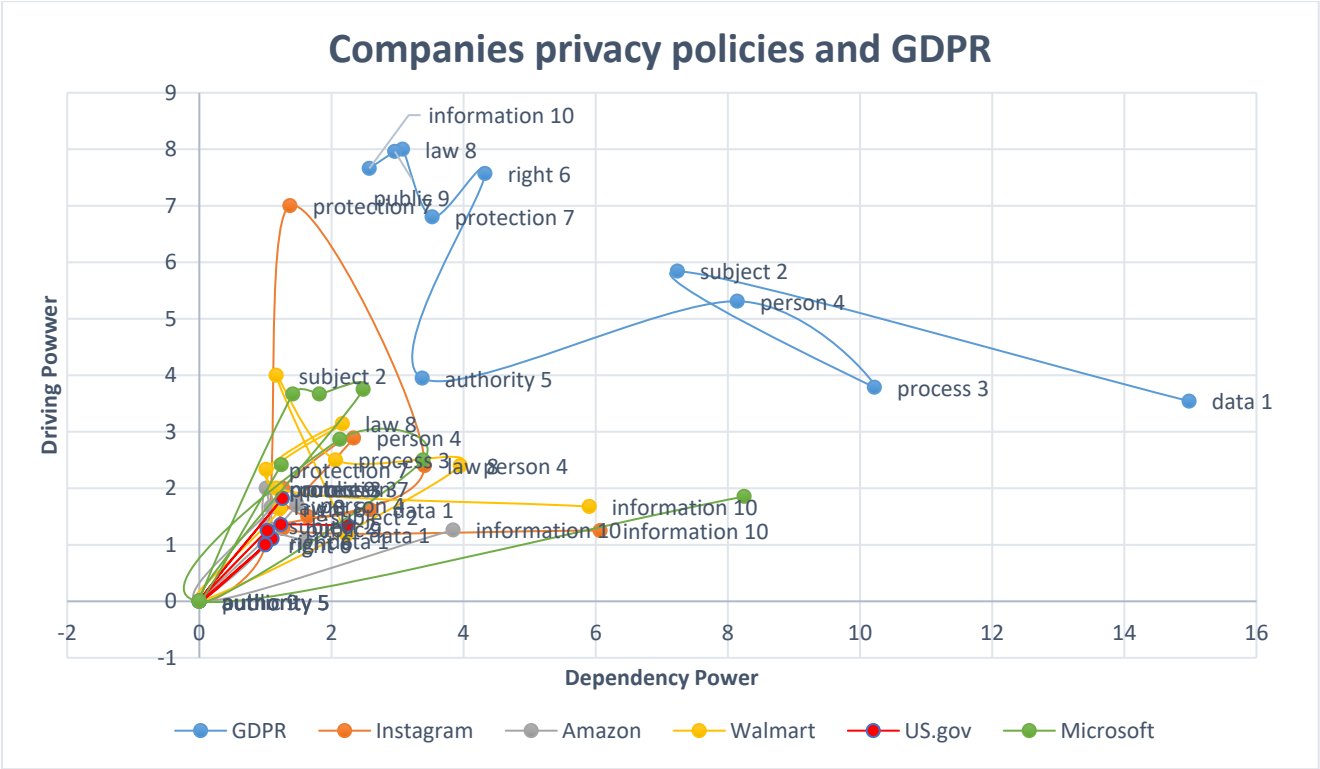

FIGURE 7: MICMAC ANALYSIS OF FIVE COMPANIES AND THE PATH OF TERMS

1) GDPR

As described in section III, 10 frequent terms from reference documents have been selected that cover the concept of the privacy policy. These terms were scored from 1 to 10 based on their
frequencies in the GDPR document. (1 is the most frequent term). Figure 6 is shown the MICMAC analysis of 5 different companies and GDPR. The line the connect the terms for each company followed the frequency order in GDPR. The path created by the connections of the terms visualized the differences in the frequency of terms in the companies. Five service
providers that we considered for our study include Amazon [17], Instagram [19], Microsoft [20], the US government [21], and Walmart [18]. The blue line in Figure 7 presents the terms in the MICMAC analysis of GDPR. Dependency power and driving power of each word cause that the terms spread all over the plot. Compared to the other five providers, lack of power for each term put the big portion of the path in the autonomous

section. Amazon and Walmart's privacy policies are analyzed separately in the next step.

Another tool that helps us to understand the differences between datasets is the hierarchical diagram. The hierarchical analysis is one of the technics from the ISM method that is customized for TFISM. The process of the hierarchical analysis is presented in our previous work in detail [14]. Figure 8 presents the hierarchical analysis of GDPR. The hierarchical plot is read from bottom to top.
GDPR factors can be put in four layers. The first layer includes "information". In the hierarchical diagram, the lowest layer (first layer) has the highest driving power and uses it to push (Change) other factors to affect the systems. The highest layer (fourth layer) has the highest dependency power and drives by other factors. So, we can say that in the GDPR document, "data", "Subject", "process", "person", and "right" are the most dependent factors that change on other factors will changes these factors. Another point about the hierarchical diagram of GDPR is that it covers all the terms. In the next parts, we can see that the structure of the document can cause the hierarchical diagram to not relate all the factors to each other.

2) Analysis of Amazon's privacy policy

In this section, we present the results of the study of the factors in Amazon's privacy policy in comparison with GDPR.

Figure 9 illustrates the MICMAC analysis of Amazon's privacy policy. Many of the terms in Amazon's privacy policy have a dependency power of 1 or close to 1, but their driving power is different. Therefore, we can see a list of terms in a vertical line with the dependency of 1 that causes density in the small part of the plot.

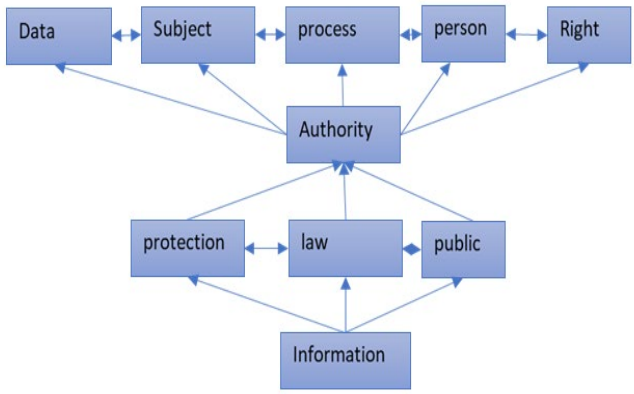

FIGURE 8:GDPR HIERARCHICAL ANALYSIS

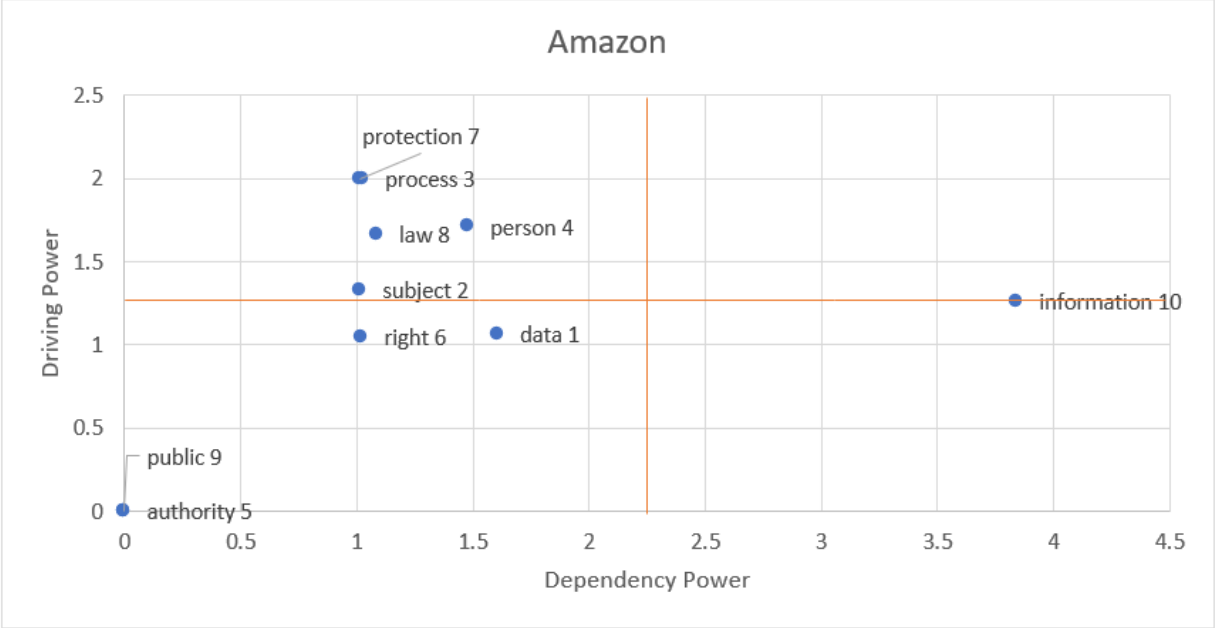

FIGURE 9: MICMAC ANALYSIS OF AMAZON PRIVACY POLICY

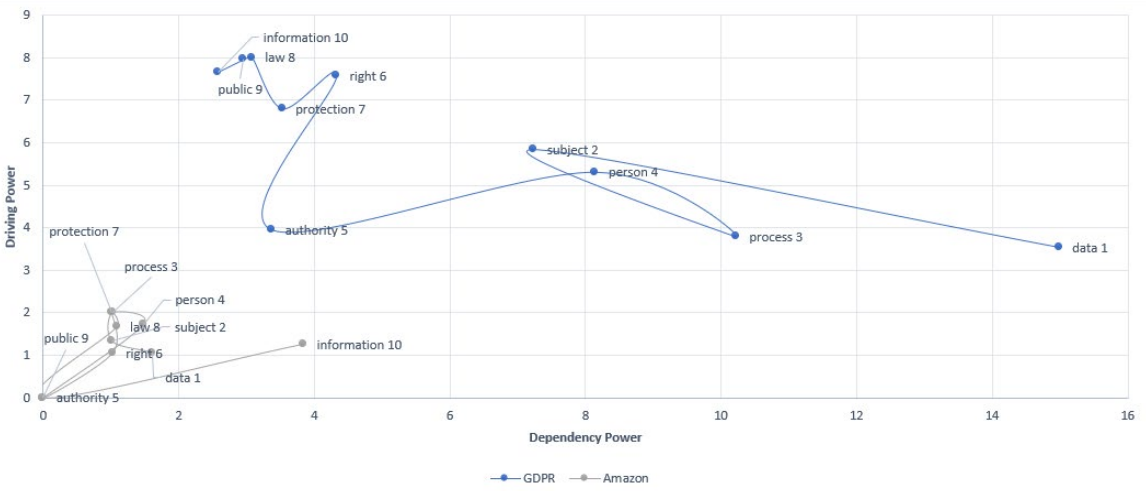

FIGURE 10: MICMAC ANALYSIS- AMAZON'S PRIVACY POLICY COMPARE TO GDPR

The MICMAC analysis of factors in the GDPR and Amazon's privacy policies are shown in Figure 10, and the dispersion and density of the plot can compare easily. One of the highlighted points in this comparison is the term "*information*". In the GDPR, information has the highest driving power and is categorized into the independent section. On the other hand, the privacy policy of Amazon's *information* has the highest dependency power and is categorized into the dependent section. This type of difference can show the different points of view about the specific factor for developing the policies.

In Figure 11 we can see the hierarchical analysis of factors in Amazon's privacy policy. This plot has four layers and the term "*right*" is in the fourth layer, but it does not have a direct connection with any other terms, and this is the reason that it is shown with yellow color. Moreover, two terms presented in orange color did not have any value to appear in the hierarchical diagram and just show for being aware of them. In GDPR layers are sequentially connected while in Amazon's privacy policy we can see that "*law*" from the first layer connects to the information in the third layer.

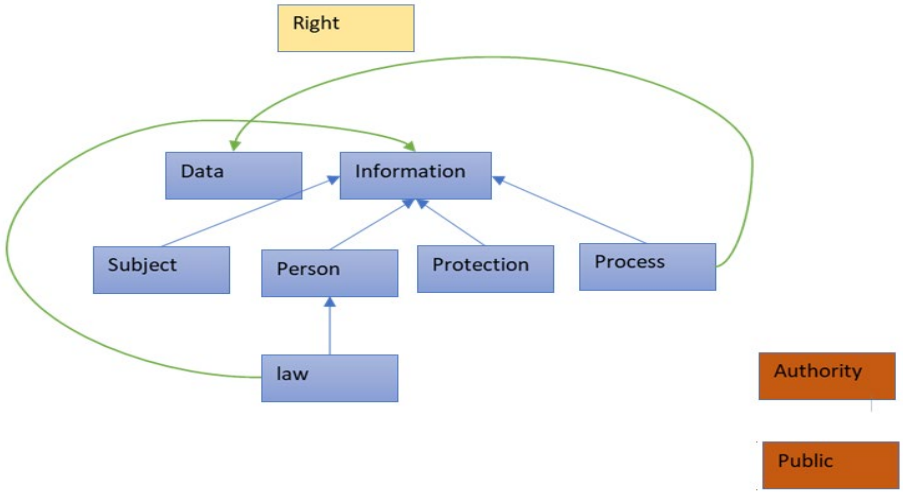

FIGURE 11: HIERARCHICAL DIAGRAM OF AMAZON PRIVACY POLICY

3) Analysis of Walmart Privacy Policy

Walmart was chosen as a second company to compare the privacy policy because in many fields Walmart and Amazon are competing and they compare with each other [22]. Figure 12 presents the MICMAC analysis of Walmart's privacy policy. In this plot "Person" is in the third section and categorize into linkage factors. This fact means in Walmart's privacy policy "person" has a critical role that can affect other factors and also any changes for "person" can resonance and change the whole system. All other variables in Walmart's privacy policy have the same category in Amazon. MICMAC analysis of GDPR, Amazon, and Walmart are shown in the one plot in Figure 13. Although all the terms in Walmart and Amazon data have lower

power than GDPR, the yellow line that belongs to Walmart is closer to the path that GDPR draws.

It is important to consider that all the values from the TFISM method are normalized and the differences between the power of the terms in GDPR and Amazon or Walmart arising from the lower frequency of these words in these documents.

Figure 14 presents the hierarchical diagram of Walmart's privacy policy. It has three levels while the GDPR has four levels. Also, "authority" did not appear in this plot. Walmart's hierarchical diagram has stronger inter-relationships between variables in each level comparing to Amazon's hierarchical diagram. Also, variables between different levels have more connections. We can say that the hierarchical diagram of Walmart's privacy policy has more similarities than the same plot from Amazon's privacy policy.

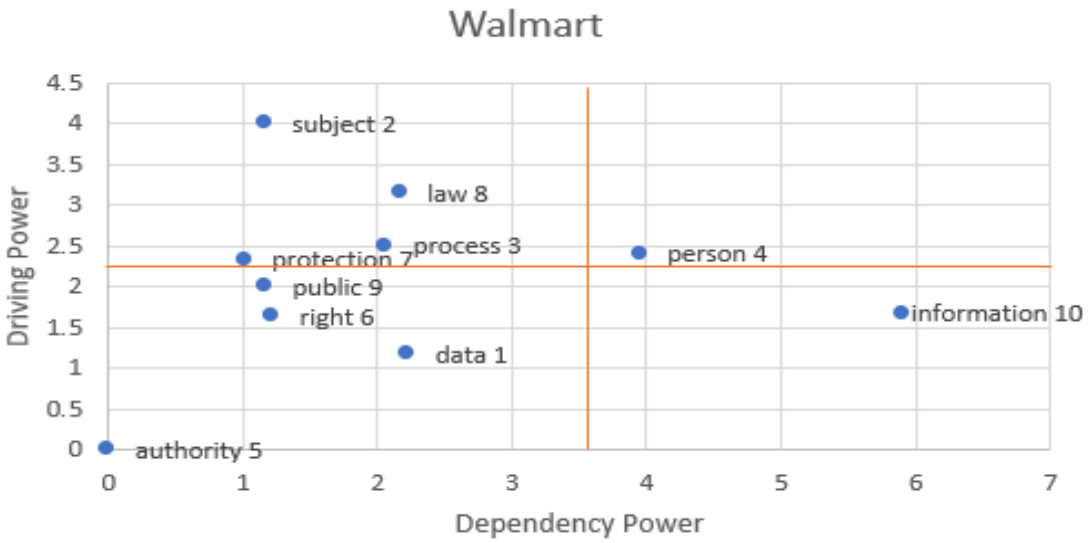

FIGURE 12: MICMAC ANALYSIS OF WALMART'S PRIVACY POLICY

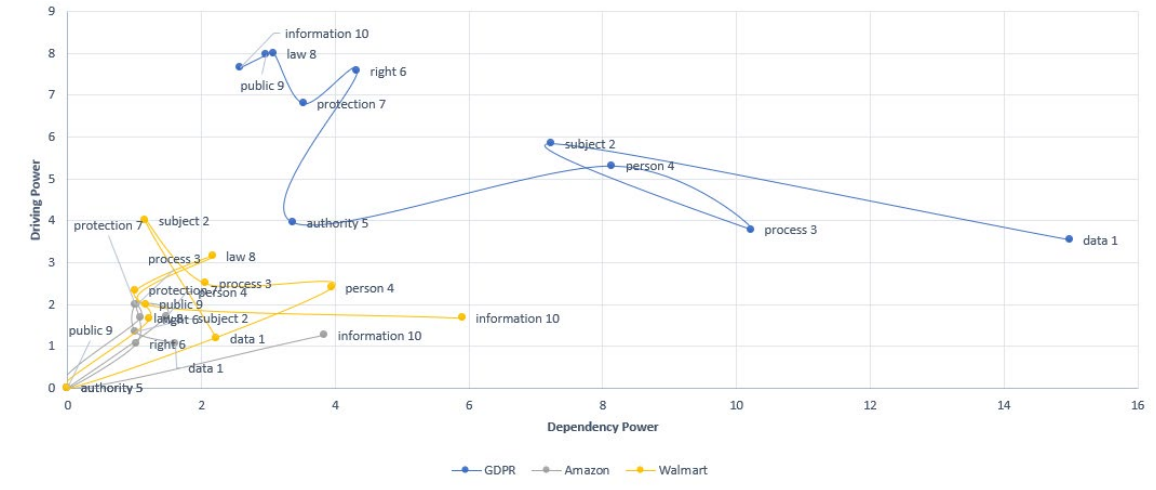

FIGURE 13: MICMAC ANALYSIS- WALMART'S PRIVACY POLICY COMPARE TO GDPR AND AMAZON'S PRIVACY POLICY

V. Conclusion

In this paper, we have presented the results of comparing the factors in the GDPR document with the same factors in cloud service privacy policies. We used a novel methodology of text analysis called Textual Fuzzy Interpretive Structural Modeling (TFISM) to illustrate the potential usage in the field of web security and privacy policy. We reference General Data Protection Regulation (GDPR) as a popular standard regulation for other policies compare with. We analyze a sample of 224 privacy policies of US companies. These companies are from various domains.

In TFISM analysis we can observe more similarities between GDPR with a combination of privacy policies. 224 selected companies come from different domains so they can cover different aspects of users' privacy policy. When we merge all the collected data consequently the merged document can cover more than every single document and it has more similarity to GDPR. Therefore, it can be concluded that the business domain is important for privacy policy coverage and for evaluating the similarity between GDPR and the company's privacy policy having subsection of domain improve the results.

TFISM functions for the cloud service privacy policy validated with the help of domain experts and, we get advantage other techniques such as association rule mining for validation the identified relationship, and statistical analysis to be confident about the factor coverage. Our corpus of the privacy policy of companies is expanding. The next phase of our research will be focused on categorized data based on the business domain.

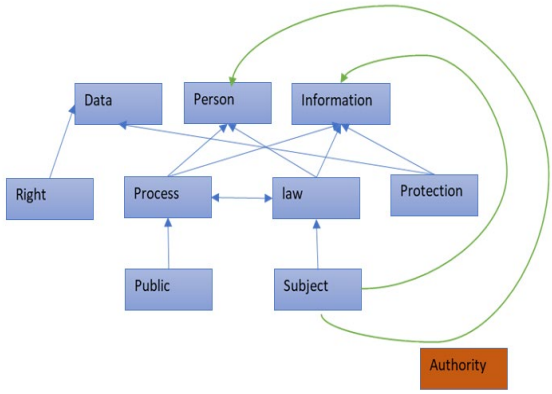

Figure 14: Hierarchical Diagram of Walmart Privacy Policy

Acknowledgment

This research was partially supported by a DoD supplement to the NSF award 1747724, Phase I IUCRC UMBC: Center for Accelerated Real-time Analytics (CARTA). We also thank Mrs. Lavanya Elluri for her professional consultation to select the list of variables and analyzing the results.

References

[1] Parliament, "https://gdpr-info.eu/," 2016. [Online]. Available: https://gdpr-info.eu/. [Accessed 2018].

[2] J. Warfield, " History and Applications of Interpretive Structural Modeling," " (1979).," May 23, 1979..

[3] R. K. RAGADE, "Fuzzy interpretive structural modeling." Cybernetics and System 6.3-4 (1976): 189-211.

[4] k. Zadeh, "Fuzzy sets, fuzzy logic, and fuzzy systems," World Scientific., p. 6, 1996.

[5] G. Khatwani, "Fuzzy-TISM: A fuzzy extension of TISM for group decision making.," Global Journal of Flexible Systems Management, pp. 97-112., 2015.

[6] S. M. T. a. F. A. F. Chandramowli, "Analysis of barriers to development in landfill communities using interpretive structural modeling.," Habitat International, pp. 246-253, 2011.

[7] J. P. a. P. V. Saxena, " Impact of indirect relationships in the classification of variables—a Micmac analysis for energy conservation.," Systems Research, pp. 245-253, 1990.

[8] S. Sushil, "Interpreting the interpretive structural model.," Global Journal of Flexible Systems Management, pp. 87-106, 2012.

[9] T. e. a. Linden, "The privacy policy landscape after the GDPR.," arXiv preprint arXiv, 201

[10] R. N. a. K. S. B. Zaeem, "The effect of the GDPR on privacy policies: recent progress and future promise.," ACM Transactions on Management Information Systems (TMIS), pp. 1-20, 2020.

[11] e. a. Chang, "Automated and personalized privacy policy extraction under GDPR consideration.," in Automated and personalized privacy policy extraction under GDPR consideration., Cham, 2019.

[12] W. B. e. a. ". Tesfay, "I read but don't agree: Privacy policy benchmarking using machine learning and the eu gdpr.," in Companion Proceedings of The Web Conference, 2018.

[13] e. a. Torre, "An ai-assisted approach for checking the completeness of privacy policies against gdpr.," in 2020 IEEE 28th International Requirements Engineering Conference (RE)., 2020.

[14] Ronak Razavisousan and Karuna P. Joshi, " Building_Textual_Fuzzy_Interpretive_Structural_Modeling_to_Analyze _Factors_of_Student_Mobility ", UMBC Technical Report, https://ebiquity.umbc.edu/paper/html/id/977, April 2021

[15] L. E. a. K. P. Joshi, "A Knowledge Representation of Cloud Data controls for EU GDPR Compliance," in IEEE World Congress on Services (SERVICES), San Francisco, 2018.

[16] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet allocation." the Journal of Machine Learning Research 3 (2003): 993-1022.

[17] Amazon, "Amazon Privacy Policy," Amazon, April 2020. [Online]. Available: https://www.amazon.com/gp/help/customer/display.html?nodeId=GX7NJQ4ZB8MHFRNJ.

[18] Walmart, "Walmart-privacy-policy," Walmart, April 2020. [Online]. Available: https://corporate.walmart.com/privacy-security/walmart-privacy-policy

[19] Instagram, "Instagram.com," Instagram, April 2020. [Online]. Available: https://help.instagram.com/519522125107875.

[20] Microsoft, "Microsoft.com," Microsoft, April 2020. [Online]. Available: https://privacy.microsoft.com/en-us/privacystatement.

[21] U. government, "https://www.usa.gov/policies," US government, April 2020. [Online]. Available: https://www.usa.gov/policies.

[22] J.-M. François, "Amazon Vs. Walmart: The Next Decade Will Decide Which Comes Out On Top," 15 01 2021. [Online]. Available: https://www.forbes.com/sites/jeanmarcfrancois/2020/01/15/amazon-vs-walmart-the-next-decade-will-decide-which-comes-out-on-top/?sh=42d341974403.