FACTA UNIVERSITATIS
Series: **Automatic Control and Robotics** Vol. 24, Nº 1, 2025, pp. 63 - 73
https://doi.org/10.22190/FUACR250429005F

Regular Paper

REDUCED CONCEPT LATTICE USABILITY – AN ANALYSIS OF METADATA EVOLVEMENT IMPACT

UDC ((004.233.3+004.4):66.094.2)

Milena Frtunić Gligorijević, Miloš Bogdanović, Darko Puflović, Leonid Stoimenov

University of Niš, Faculty of Electronic Engineering,
Computer Science Department, Republic of Serbia

ORCID iDs: Milena Frtunić Gligorijević https://orcid.org/0000-0002-9779-5465
Miloš Bogdanović https://orcid.org/0000-0002-9262-2821
Darko Puflović https://orcid.org/0000-0002-4406-3420
Leonid Stoimenov https://orcid.org/0000-0003-1097-9833

Abstract. *Numerous datasets have been made available on open data portals as a result of the open data initiatives. These portals offer a variety of search possibilities based on the metadata of datasets to facilitate data findability and usability. However, insufficient information frequently has a direct effect on search result quality and, in turn, data discoverability. As a result, methods for completing the missing metadata information—such as missing dataset category values—have become necessary. One of these methods focuses on classifying datasets according to the tags that are applied to them. The foundation of this method is a knowledge base made up of concept lattices produced for every category using the Formal Concept Analysis method. We analyze two sets of reduced concept lattices created for Ireland's open data portal datasets in 2020 and 2021 and their usability for categorizing new datasets that were available on the portal in 2021 and 2023. Among other results, we will show that concept lattices, although reduced, can be used for a long period and still preserve the accuracy of the categorization algorithm above 90%.*

Key words: *Open data portal, categorization, formal concept analysis, concept lattice, reduction.*

Received April 29, 2025 / Accepted June 09, 2025

Corresponding author: Milena Frtunić Gligorijević
University of Niš, Faculty of Electronic Engineering, Department of Computer Science, Aleksandra Medvedeva 4, 18000 Niš, Republic of Serbia
E-mail: milena.frtunic.gligorijevic@elfak.ni.ac.rs

1. INTRODUCTION

There has been a widespread and overwhelming tendency of releasing material on the Web as a result of the numerous projects in the past. The concept of open data, which involves making data publicly accessible for use, is the driving force behind these projects. It is anticipated that the public, corporate, and individual sectors would gain from open data, which is viewed as a catalyst for economic growth, knowledge, and innovation [1]. According to 2018 research funded by the European Commission, the overall value increase of public sector information is expected to reach 215 billion over the next ten years [2].

Government departments are publishing data from a variety of areas, including statistics, transportation, the environment, permits, licenses, budget, geography, and the economy, in an effort to improve transparency and inform civil societies [3]. This is known as Open Government Data (OGD). Data is released on open data portals to realize its full value. Digital libraries with catalogs and databases for file categorization are comparable to these portals. Every dataset that is made available on the Open Data Portal is accompanied by metadata, which is structured descriptive data that contains a variety of details about the resource in question. Keys include the names of properties, and values represent the information that corresponds to those properties. Metadata entries are arranged as key-value pairs.

In order to make data discoverability easier, open data portals offer search tools based on the metadata. The category and tags are the foundation of some of the simplest, most user-friendly search choices. While tag-based search offers more focused browsing based on the terms used for the dataset's description, category-based search chooses datasets that are related to the same subject. However, insufficient or inaccurate information makes it difficult for the search process to work, produces partial results, and makes data harder to find and discover. The volume of public datasets and the expansion of the open data portal are making this issue more noticeable.

The completeness of metadata entries determines the quality of search results on open data portals. Owing to the growing amount of datasets available on open data portals, insufficient metadata significantly affects the discoverability, subsequent usage, and reuse of datasets, especially when no categorization is specified. The significance of metadata for dataset discovery [4][5] and accurate interpretation [6] is suggested by a number of research studies. The scientific community has since focused much on enhancing the quality of metadata. Nevertheless, a number of studies have demonstrated that metadata is frequently lacking [7], which has prompted the development of methods for completing the gaps in the metadata.

One of the ways for filling in the missing category information is based on the existing metadata information, notably tags used for characterizing already classified datasets. The strategy depends on the Formal Concept Analysis (FCA) method for building concept lattices for each accessible category on an open data portal. These concept lattices serve as a knowledge base for classifying datasets according to the tags that characterize them and a hierarchy of tag usage within a single category. But as the number of datasets on a portal grows, so does the number of tags and tag combinations used to describe the datasets. Therefore, it is crucial to reconstruct the knowledge base or a portion of it, i.e., the concept lattices for categories that have undergone considerable change, by adding new datasets with different tags in order to preserve the accuracy of the categorization process over time.

In this paper we analyze the usability and efficiency of the reduced concept lattices for the task of the categorization of newly added dataset on open data portal. Our analysis is focused on measuring the usability of two knowledge bases by evaluating the precision of the FCA-based categorization algorithm. For this purpose we are using two datasets gathered from Ireland's open data portal in 2020 and 2021 for creating two reduced knowledge bases, by reducing the formal contexts and creating concept lattices using FCA. Using such created knowledge bases we performed 3 tests, and we analyze the categorization success and therefore the usability of reduced concept lattices. In the first test, we used 2020 knowledge base and categorized all new datasets that were available in 2021. In the second test, we used the same knowledge base for categorization of all new datasets that were available in 2023. Lastly, in the third test, we used 2021 knowledge base for categorization of all new datasets that were available in 2023.

2. Related Work

Formal concept analysis, often known as the "applied lattice theory" technique, was officially described by Rudolf Wille in 1982 [8] as a method based on the lattice theory that is used to arrange concepts in hierarchies and extract significant relationships from them. Due to the fact that no information was lost in the process, FCA quickly became popular in data analysis. Nevertheless, there is a potentially significant trade-off in computing costs. The analysis of the link between the collection of objects and their characteristics in each domain is a part of the fundamental FCA workflow, and the findings are represented as a binary matrix called formal context. To put it another way, the formal context is an object-attribute matrix, with attributes in columns and objects in rows. The cell where the row corresponding to the object and the column corresponding to the attribute overlap will have a cross if the object contains an attribute. A collection of objects and characteristics that reflect the domain knowledge are produced when the FCA theory is applied to a formal environment. These may then be arranged into a lattice diagram to extract different rules and connections.

The foundation of this method is presented by the following definitions:

Definition 1 [8] A formal context is a triple $K := (G, M, I)$ which consists of a set G of objects, a set M of attributes, and a binary relation $I \subseteq G \times M$. $(g, m) \in I$ is read as "object g has attribute m".

Definition 2 [8] For $A \subseteq G$, let $AI := \{m \in M \mid \forall g \in A: (g, m) \in I\}$, and dually, for $B \subseteq M$, let $BI := \{g \in G \mid \forall m \in B: (g, m) \in I\}$.

If the following conditions are met: $A \subseteq G$, $B \subseteq M$, $AI = B$, $BI = A$, then a pair (A, B) is a formal context. Set A is named the concept extent while set B is named the concept intent.

Definition 3 [8] The set S(C) of all concepts of a formal context C together with a partial order $(A1, B1) \leq (A2, B2) :\leftrightarrow A1 \subseteq A2$ (which is equivalent to $B1 \supseteq B2$) is a complete lattice of C.

Since the domain expert's examination of the concept lattice enables the discovery of relationships in the data, the computation of the set of all concepts and its line diagram is crucial for applications utilizing FCA. FCA has been used for knowledge discovery [9][10][11][12][13], knowledge representation [14], pattern matching problems [15][16], web usage mining [17][18] and many other areas of data science. To illustrate the literature on FCA, Poelmans et al. used FCA [9]. They conducted a study of 702 publications that were published between 2003 and 2009 and included the formal concept analysis in the abstract.

They specifically took advantage of 140 studies that were only focused on data mining and knowledge discovery. The primary research subjects in the FCA community were represented and visualized by a re-attempt of an extensive survey in 2013 [10]. From 2011 to 2016, Singh et al. [19] examined over 350 studies. The following tendencies in FCA research were found by them: FCA such as granular computing, FCA with a fuzzy setting, an interval-valued fuzzy setting, possibility theory, a rough setting, a triadic setting, factor concepts and the incomplete context. Their research confirmed FCA to be used for knowledge discovery, reasoning, decision context and ontology engineering. By surveying several FCA-based classification techniques and classifying them into three categories – methods based on distributed classifiers, ensemble classifiers, and mono-classifiers—Azibi et al. [20] demonstrated the utility of FCA in machine learning.

In the era of big data, FCA may also be used as a potent tool for large-scale dataset analysis. Therefore, having quick and precise FCA methods for knowledge representation and discovery is crucial when working with big datasets. To expedite the process of listing every potential notion (relevant concept), a number of distributed and parallel alternatives have been proposed in recent years. The study's authors suggest parallel implementations that make use of the CloseByOne algorithm [21]. In 2009, the authors developed the first distributed algorithm [23] using the MapReduce programming architecture [22]. In order to extract new information from binary object-attribute relational data, Chunduri et al. [24] presented a distributed FCA technique they called the UNConceptGeneration.

To address the issue of creating idea lattices of sufficient size and structure to exhibit crucial context features, a number of attribute and concept lattice reduction techniques are put forth in the literature; each technique has a unique set of characteristics. Some approaches begin the reduction of concept lattice complexity with a context-level representation. On the other side, we can carry out lattice level pruning, such as the iceberg reduction, for some techniques that use a lattice version to minimize complexity. In general, certain reduction techniques aim to identify the fewest objects or characteristics that maintain the structure of the original lattice during the reduction process. In general, academia has reported [25] context pro-processing methods [26][27][28][29], non-essential distinctions elimination [30][31][32] and concept filtration [33][34].

3. Formal Context Reduction

Multiple researches have stated the importance of metadata completeness for data to reach its full potential. One of the vital metadata components for this purpose is information regarding the category a dataset should belong to. However, bearing in mind that this information is often missing [7], it is important to develop a mechanism for filling out missing metadata values such as this one. Methods relying on Formal Concept Analysis for filling out missing metadata information have shown great potential for suggesting appropriate categories based on the keywords used for their description [35]. However, since such methods rely on knowledge bases, the update or the recreation of the part of the knowledge base or the whole knowledge base periodically has to be performed for this approach to preserve its precision. For tasks performed on open data portals such as filling in dataset category based on tag values, FCA-based methods can use concept lattices for representing tags usage within categories, having one concept lattice per category.

Concept lattice generation can be a demanding process and algorithms for concept lattice generation often expose an exponential dependence regarding the number of used attributes and/or objects. Therefore, a reduction of formal context [37] used as the base for performing FCA and generating concept lattices is one way of addressing this problem. The number of distinct tag values that constitute formal context is directly proportional to the concept lattice scale. Therefore, the optimization of tags usage within categories through means of semantic similarity measure can reduce the number of appearing tags and consequently affect the size of the concept lattice.

As stated in [37], the reduction algorithm starts by converting words that represent each tag in category into 300-dimensional vectors using GloVe (Global Vectors for Word Representation) [36] model. After that, similarity for each pair of tags belonging to the same category is calculated using cosine similarity according to the following definition:

Definition 4 [37] Similarity S_T between two tags, T_X and T_Y, is defined as the largest similarity between any two words W_X and W_Y, whereas W_X and W_Y represent parts of T_X and T_Y, respectively.

After tag similarities are calculated for each pair of tags, a threshold set to 0.8 is used to determine a set of similar tags. In this way, two tags are considered similar if their similarity value is higher than the set threshold. The threshold value depends on the model used for word vectorization. Although, 0.7 threshold value is considered to be appropriate for the 300-dimensional models [38], a more restricted threshold value was set for this analysis. Using this approach, each tag is coupled with a list of similar tags, and if the list of similar tags contains more than 2 tag value, tag is identified as reduction candidate tag.

In the next step, lists of similar tags are compared to determine the largest subset that is equal to an existing list. Once such list is identified, the largest subset of tags is replaced by its corresponding tag. This step repeats until all subsets have been identified and replaced with their corresponding tags. Tag replacements performed in this step reduce the number of tags that appear in a formal context, making the context more concise and simpler without losing information. Performing a FCA method on such reduced formal context, a reduced concept lattices are generated.

4. Analysis of the Usability of Reduced Formal Concept Lattices

Within this research, we are focused on analyzing the usability of concept lattices used for the categorization of datasets on an open data portal based on their tags. Our analysis is confined to the reduced concept lattices, i.e. concept lattices where similar tags are replaced by a single value to simplify the concept lattices. Within our analysis, we are analyzing the evolution of datasets on the open data portal and consequently, changes in concept lattices for its categories. Moreover, we analyze the categorization success for new datasets added over time on the portal, based on the created reduced concept lattices using the approach for metadata enrichment presented in the research [35].

For the purpose of the analysis of the usability of the reduced concept lattices over time, we have used 3 sets of data from Ireland open data portal:

- dataset 2020 - set containing all datasets that were available on the portal in 2020,
- dataset 2021 - set containing all datasets that were available on the portal in 2021,
- dataset 2023 set containing all datasets that were available on the portal in 2023.

Ireland open data portal organizes its data into 14 categories, whereas each dataset belongs to a single category. Further, for this portal, an upward trend in the number of datasets was noted in the period analyzed. More specifically, in 2020, the total number of datasets was 10151, in 2021 there were 12949 available datasets, and in 2023 total number of datasets was 17844. Furthermore, the total number of tags was 5070 in 2020, 11778 in 2021, and 15593 in 2023. However, it was noticed that although the total number of datasets increased by less than 3000, in 2021 there were 7984 datasets with identification numbers (ids) that were not on the portal in 2020. Consequently, the distribution of datasets by category was significantly changed. The number of datasets was significantly reduced in some categories, like *Economy* and *Society*, and considerably increased in categories like *Government*. The detailed distribution of the number of available datasets by category for both years is presented in Table 1. Furthermore, such changes in the available content affected the usage of tags and combination of tags, and it was noticed that the average percentage of new tags for all categories in 2021 was 49.17%.

Datasets 2020 and 2021 were used for creating reduced knowledge bases 2020 and 2021, respectively. These knowledge bases were used for evaluating reduced concept lattice usability over time and were generated the using formal context reduction method presented in the previous section of this research and FCA. In other words, for both knowledge bases, the formal context for each category on the portal was reduced, and using such reduced formal context, concept lattices were generated. The results of the reduction algorithm for each category on the portal for both years are presented in Table 1.

Table 1 Reduction information for formal contexts 2020 and 2021

Category	2020				2021			
	ND	NT-BR	NT-AF	NT%	ND	NT-BR	NT-AF	NT%
Agriculture	123	105	96	8.57	131	270	259	4.07
Arts	135	297	290	2.36	65	215	207	3.72
Crime	8	13	13	0.00	93	100	100	0.00
Economy	1149	202	191	5.45	321	416	398	4.33
Education and Sport	182	154	149	3.25	225	364	357	1.92
Energy	123	243	226	7.00	125	243	230	5.35
Environment	3098	1712	1575	8.00	3444	2144	1973	7.98
Government	495	1409	1311	6.96	5775	6772	6672	1.48
Health	1416	427	406	4.92	1817	971	951	2.06
Housing	424	347	335	3.46	345	388	373	3.87
Science	329	387	375	3.10	127	376	355	5.59
Society	2203	459	442	3.70	86	195	169	13.33
Towns	50	52	52	0.00	18	34	34	0.00
Transport	411	330	319	3.33	370	542	520	4.06

ND – number of datasets
NT-BR – number of tags before reduction
NT-AF – number of tags after reduction
NT% – percentage reduction in number of tags

From the presented reduction information, it can be noted that only 2 categories, *Crime* and *Towns,* did not have any reduction in the number of tags for both years. However, these categories are also the smallest in terms of the number of tags appearing within them. The number of tags in other categories was reduced, with the largest reduction recorded in the

2021 dataset in the category *Society*, 13.33%. The structures of concept lattices, which constitute knowledge bases 2020 and 2021, are presented in Table 2. Within the presented table, for each concept lattice, the total number of nodes and the number of levels is presented.

Table 2 Structure of concept lattices for knowledge bases 2020 and 2021

Category	2020		2021	
	Number of nodes	Number of levels	Number of nodes	Number of levels
Agriculture	35	6	154	6
Arts	251	8	133	9
Crime	6	3	96	3
Economy	123	6	326	8
Education and Sport	86	5	259	5
Energy	297	15	224	12
Environment	3996	21	5080	21
Government	977	12	6257	13
Health	219	7	773	9
Housing	490	9	553	9
Science	358	17	450	17
Society	376	7	90	7
Towns	31	4	22	3
Transport	256	7	483	7

From the presented data it can be noticed that some of the concept lattices have significantly different structure in 2021 compared to 2020. These differences are due to the sizable changes that occurred in the period from 2020 to 2021 in terms of available datasets within categories and tags and combination of tags used within them.

These knowledge bases, e.g. concept lattices, were used for 3 test cases. Knowledge base 2020 was used for categorization of all new datasets that were available in 2021 (2020-2021 test dataset), that is, datasets from 2021 whose identifiers were not in the 2020 dataset. Further, this knowledge base was used for categorization of all new datasets that were available in 2023 (2020-2023 test dataset), that is, datasets from 2023 whose identifiers were not in the 2020 dataset. Further, knowledge base 2021 was used for categorization of all new datasets that were available in 2023 (2021-2023 test dataset), that is, datasets from 2023 whose identifiers were not in the 2021 dataset. Therefore, the distribution of the number of examples in test cases is different and the total number of datasets in each test set is presented in Table 3 (column ND). Also, the distribution of test datasets by categories differs, since it depends on datasets that have been added to the portal in the meantime.

The categorization results for all 3 test cases are presented in Table 3. Since datasets on Ireland open data portal are assigned to at most one category, and used categorization algorithm is designed to suggest multiple categories if the criterion is met, all categorization results were split into 3 groups. First group, *FM*, presents datasets that were fully categorized, meaning that they were assigned a correct category with no additional category suggested. The second group, *FME*, presents datasets that were assigned a correct category and at least one additional category. Last group, *MM*, presents mismatched datasets – datasets that were assigned incorrect category. Since for both *FM* and *FME* groups, categorization algorithm assigns correct categories, datasets belonging to these groups will be considered correctly categorized.

Table 3 Categorization results

Test dataset	ND	FM [%]	FME [%]	MM [%]
2020-2021	7984	43	3	54
2020-2023	12979	46	7	47
2021-2023	5541	70	14	15

ND – number of datasets
FM [%] – percentage of fully matched
FME [%] – percentage of fully matched datasets with extra categories
MM [%] – percentage of mismatched datasets

As can be seen from the presented results, in the first test dataset (test dataset 2020-2021), 46% of datasets belonged to *FM* and *FME* groups, meaning that the algorithm assigned a correct category to a dataset. Although the total percentage of successfully categorized datasets is below 50%, test datasets belonging to category *Transport* were successfully categorized in 99.61% of the test cases, and test datasets belonging to category *Energy* were successfully categorized in 94.92% of the test cases. Further, the percentage of successfully categorized datasets was above 70% for categories *Science* and *Environment,* and above 85% for category *Arts*. The lowest percentage of successfully categorized datasets was obtained for categories *Crime*, *Towns*, and *Education and Sport*, where the percentage was below 10%.

Additionally, the results of categorization of the second test set (test dataset 2020-2023) showed that 53% of test datasets belonged to groups *FM* and *FME*. Analyzing categories separately, in this test set, only 3 categories had percentage of the successfully categorized datasets above 70%. Datasets belonging to the category *Environment* were successfully categorized in 77.51% of cases, while datasets belonging to category *Arts* were successfully categorized in 88.37% of cases. The highest categorization results were obtained for category *Transport*, where 98,01% of datasets were successfully categorized. The lowest percentage of successfully categorized datasets was obtained for category *Crime* where the percentage was below 0,49%. Other than this category, only category *Towns* had the percentage of successfully categorized datasets below 10%.

The best categorization results were obtained for the third test case (test dataset 2021-2023), where 84% of datasets belonged to groups *FM* and *FME*. In this test, analyzed by categories, datasets that belonged to categories *Health*, *Government*, *Arts*, *Crime*, and *Transport* had more than 92% of successfully categorized datasets, and category *Towns* was only category with the percentage of successfully categorized datasets below 10%.

Based on this analysis, it can be concluded that knowledge base 2020 had significantly lower categorization results compared to knowledge base 2021. This result was expected, due to the major changes made on the portal in the period from 2020 to 2021. Furthermore, based on the results for knowledge base 2021, and some of the categories from knowledge base 2020 that were not significantly changed in the period from 2020 to 2021, it can be concluded that reduced concept lattices can remain a solid knowledge base for a long period without the need for their update.

4. Conclusion

Approaches based on the Formal Concept Analysis have shown a great potential for filling in missing metadata values, particularly information regarding a category a dataset should belong to. However, for task of categorization of datasets based of their tags, this approach relies on the knowledge base that contains one concept lattice for each category available on open data portal. For the categorization method to preserve its accuracy these knowledge bases require periodical update, depending on changes on the portal and its growth. Furthermore, concept lattices grow with the increase of the number of datasets, tags, and combinations combination of tags, with tendency to become too complex. Therefore, a method for reduction of formal contexts can be used for generating reduced formal contexts and consequently reduced concept lattices. These concept lattices reduce the heterogeneity of tag values while preserving the meaning of tags for a particular open dataset category.

Within this paper, we have analyzed the usability of reduced concept lattices over time for categorization of datasets on open data portals. We have used data from Ireland open data portal, and we compared 2 sets of concept lattices created based on available data in 2020 and 2021. Further, we analyzed its usability and efficiency for categorization of a new dataset through 3 test cases. Significant updates were made to the portal during the analyzed period, which influenced the availability of datasets, the total number of datasets, the variety of tags, and combination of tags appearing within categories. Therefore, these changes had an impact on the structure of formal contexts and concept lattices.

As a result of such significant changes made to the portal, the categorization of new datasets in 2021 using 2020 knowledge base was successful in less than 50% of test cases. This percentage was equal to 53% for categorization of new datasets in 2023 using 2020 knowledge base. The results obtained for categorization of new datasets in 2023 using knowledge base 2021 were much higher, with 84% of datasets correctly categorized. Therefore, based on the results of using the 2021 knowledge base and individual concept lattices that did not change significantly between 2020 and 2021, it can be concluded that reduced lattices can remain a solid knowledge base for a long period. However, if major changes are made on the open data portal or in part of the categories, the knowledge base should be recreated or updated, for the categorization algorithm to preserve its precision. Therefore, it can be concluded that concept lattices reduced using this approach show a great potential for servings as knowledge base for categorization purposes on portal such as Ireland ODP where datasets are categorized in only one category.

In the future, this analysis should be performed on data provided by portals where datasets are categorized with multiple categories. Such additional analysis will provide a deeper insight into the usability of reduced concept lattices on portals where tags are differently distributed among categories. Furthermore, within this analysis, we used a 0.8 threshold value for reduction as a more restrictive value. In the future, different threshold values can be applied in the reduction phase, and differences in the performance of concept lattices created in that way can be examined.

Acknowledgement: *This work was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia [grant number 451-03-137/2025-03/200102].*

REFERENCES

[1] J. Berends, W. Carrara and C. Radu, "The economic benefits of open data," Analytical report 9, 2017.

[2] M. Barbero et al., "Study to support the review of directive 2003/98. EC on the re-use of public sector information," 2018, ISBN 978-92-79-83169-0. [Online]. Available: http://doi.org/10.2759/ 373622.

[3] T. Davies, S.B. Walker, M. Rubinstei and F. Perini, "The State of Open Data: Histories and Horizons," Cape Town and Ottawa: African Minds and IDRC, Published by African Minds and the IDRC, 2019. [Online]. Available: http://doi.org/10.5281/zenodo.2668475.

[4] K. Braunschweig, J. Eberius, M. Thiele, et al., "The State of Open Data Limits of Current Open Data Platforms", 2012.

[5] K. J. Reiche, E. Höfig, "Implementation of metadata quality metrics and application on public government data," In *2013 IEEE 37th An-nual Computer Software and Applications Conference Workshops*, pp. 236–241, 2013. [Online]. Available: http://doi.org/10.1109/COMPSACW.2013. 32.

[6] A. Zuiderwijk, C. Volten, M. Kroesen, et al, "Motivation perspectives on opening up municipality data: Does municipality size matter?" Information 9, no. 11: 267. [Online]. Available: http://doi.org/10.3390/info9110267.

[7] M. Frtunić Gligorijević, M. Bogdanović, and L. Stoimenov, "Tracking metadata changes in the government open data portals," In *ICIST 2022 Proceedings*, pp.180-184, 2022. ISBN 978-86-85525-24-7

[8] R. Wille, "Restructuring lattice theory: an approach based on hierarchies of concepts," In: *Ferré S., Rudolph S. (eds) Formal Concept Analysis. Lecture Notes in Computer Science*. Vol. 5548, Springer, Berlin, Heidelberg, 2009.

[9] J. Poelmans, P. Elzinga, S. Viaene, G. Dedene, "Formal concept analysis in knowledge discovery: a survey," In *Proceedings of the International conference on conceptual structures*, pp. 139-153. Springer, Berlin, Heidelberg, 2009.

[10] J. Poelmans, S. O. Kuznetsov, D. I. Ignatov, D. Dedene, "Formal concept analysis in knowledge processing: A survey on models and techniques," Expert systems with applications. 40(16), 6601-6623, 2013.

[11] M. W. Chekol, A. Napoli, "An FCA framework for knowledge discovery in SPARQL query answers," In *Proceedings of the 12th International Semantic Web Conference*, 2013.

[12] P. Valtchev, R. Missaoui, R. Godin, "Formal concept analysis for knowledge discovery and data mining: The new challenges," In *Proceedings of the International conference on formal concept analysis,* pp. 352-371, Springer, Berlin, Heidelberg, 2004.

[13] M. Alam, A. Buzmakov, V. Codocedo, A. Napoli, "Mining definitions from RDF annotations using formal concept analysis," In *Proceedings of the International Joint Conference in Artificial Intelligence*, 2015.

[14] M. Alam, T. N. N. Le, A. Napoli, "Latviz: A new practical tool for performing interactive exploration over concept lattices," In *CLA 2016-Thirteenth International Conference on Concept Lattices and Their Applications*, 2016.

[15] F. Venter, D. G. Kourie, B. W. Watson, "FCA-based two dimensional pattern matching," In *International Conference on Formal Concept Analysis*, pp. 299-313. Springer, Berlin, Heidelberg, 2009.

[16] G. Li, "DeepFCA: Matching biomedical ontologies using formal concept analysis embedding techniques," In *Proceedings of the 4th International Conference on Medical and Health Informatics*, pp. 259-265, 2020.

[17] B. Zhou, S. C. Hui, K. Chang, "A formal concept analysis approach for web usage mining," In *Proceedings of the International Conference on Intelligent Information Processing*, pp. 437-441, Springer, Boston, MA, 2004.

[18] H. He, H. Hai, W. Rujing, "FCA-based web user profile mining for topics of interest," In *Proceedings of the IEEE International Conference on Integration Technology*, pp. 778-782, 2007.

[19] P. K. Singh, A. K. Cherukuri, A. Gani, "A comprehensive survey on formal concept analysis, its research trends and applications," International Journal of Applied Mathematics and Computer Science, 2016. [Online]. Available: http://doi.org/10.1515/amcs-2016-0035

[20] H. Azibi, N. Meddouri, M. Maddouri, "Survey on Formal Concept Analysis Based Supervised Classification Techniques, Machine Learning and Artificial Intelligence, pp. 21-29, IOS Press, 2020. [Online]. Available: http://doi.org/10.3233/FAIA200762

[21] P. Krajca, J. Outrata, V. Vychodil, "Parallel recursive algorithm for FCA," in CLA, pp. 71–82, 2008.

[22] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, pp. 137-150, 2004.

[23] P. Krajcaand, V. Vychodil, "Distributed algorithm for computing formal concepts using map-reduce framework," in International Symposium on Intelligent Data Analysis,pp. 333–344, 2009. [Online]. Available: http://doi.org/10.1007/978-3-642-03915-7_29.

[24] R. K. Chunduri, A. K. Cherukuri, M. Tamir, "Concept generation in formal concept analysis using MapReduce framework," in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, pp. 191–204, 2017. [Online]. Available: http://doi.org/10.1109/ICBDACI.2017.8070834.

[25] M. Alwersh, L. Kovács, “Survey on attribute and concept reduction methods in formal concept analysis,” Indonesian Journal of Electrical Engineering and Computer Science, vol.30, no.1, pp. 366~387, 2023. [Online]. Available: http://doi.org/10.11591/ijeecs.v30.i1.pp366-387

[26] J. Li, C. Mei, Y. Lv, “A heuristic knowledge-reduction method for decision formal contexts,” Computers & Mathematics with Applications, vol. 61, no. 4, pp. 1096–1106, 2011. [Online]. Available: http://doi.org/10.1016/j.camwa.2010.12.060.

[27] J. Medina, “Relating attribute reduction in formal, object-oriented and property-oriented concept lattices,” Computers & Mathematics with Applications, vol.64, no. 6, pp. 1992–2002, 2012. [Online]. Available: http://doi.org/10.1016/j.camwa.2012.03.087.

[28] H. Wang W.-X. Zhang, “Approaches to knowledge reduction in generalized consistent decision formal context,” Math Comput Model, vol. 48, no. 11–12, pp. 1677–1684, 2008. doi: 10.1016/j.mcm.2008.06.007.

[29] L. Antoni, M. E. Cornejo, J. Medina, E. Ramírez-Poussa, “Attribute classification and reduct computation in multi-adjoint concept lattices,” IEEE Transactions on Fuzzy Systems, vol. 29, no. 5, pp. 1121–1132, 2020. [Online]. Available: http://doi.org/10.1109/TFUZZ.2020.2969114.

[30] S. M. Dias N. Vieira, “Reducing the size of concept lattices: The JBOS approach,” in Cla, vol. 672, pp. 80–91, 2010.

[31] V. Codocedo, C. Taramasco, H. Astudillo, “Cheating to achieve formal concept analysis over a large formal context,” in *The Eighth International Conference on Concept Lattices and their Applications-CLA 2011*, pp. 349–362, 2011.

[32] C. A. Kumar, S. Srinivas, “Mining associations in health care data using formal concept analysis and singular value decomposition,” J Biol Syst, vol. 18, no. 04, pp. 787–807, 2010. [Online]. Available: http://doi.org/10.1142/S0218339010003512.

[33] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, L. Lakhal, “Computing iceberg concept lattices with titanic,” Data & Knowledge Engineering, vol. 42, no. 2, pp. 189–222, 2002. [Online]. Available: http://doi.org/10.1016/S0169-023X(02)00057-5.

[34] P. K. Singh, C. A. Kumar, “Concept lattice reduction using different subset of attributes as information granules,” Granular Computing, vol. 2, no. 3, pp. 159–173, 2017. [Online]. Available: http://doi.org/10.1007/s41066-016-0036-z.

[35] M. Frtunić Gligorijević, M. Bogdanović, L. Stoimenov, "An Approach for Metadata Enrichment on Open Data Portals," In *11th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN)*, Nis, Serbia, pp. 1-5, 2024.

[36] J. Pennington, R. Socher, C. Manning, “Glove: Global vectors for word representation,” In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532-1543, 2014.

[37] M. Bogdanović, N. Veljković, M. Frtunić Gligorijević, D. Puflović, L. Stoimenov, “On revealing shared conceptualization among open datasets,” *Journal of Web Semantics*, vol. 66, 2021. [Online]. Available: http://doi.org/10.1016/j.websem.2020.100624.

[38] N. Rekabsaz, M. Lupu A. Hanbury, “Exploration of a Threshold for Similarity based on Uncertainty in Word Embedding,” In: Jose, J., et al. Advances in Information Retrieval. ECIR 2017. Lecture Notes in Computer Science(), vol 10193. Springer, Cham., 2017. [Online]. Available: http://doi.org/10.1007/978-3-319-56608-5_31