Published Online 1-03-2021

TAXONOMY OF DATA INCONSISTENCIES IN BIG DATA

Ms. Vinaya Keskar[1], Dr. Jyoti Yadav[2], Dr. Ajay Kumar[3]

[1]Research Scholar at Savitribai Phule Pune University, Pune

Asst. Professor, ATSS's College of Business Studies and Computer Applications, Pune.

vasanti.keskar@gmail.com
[2]Department of Computer Science, Savitribai Phule Pune University, 411007, India
yadav.jyo@gmail.com
[3]Director, JSPM's Jaywant Technical Campus, Pune
ajay19_61@rediffmail.com

Abstract

In the coming years, common units of measuring data viz. kilobytes, megabytes, gigabytes, or even terabytes will begin to appear quainter as the entire digital universe is expected to produce approximately 463 Exabyte's of data every 24 hours worldwide. This omnipresent data is potentially knowledge-rich. Unprocessed data can be excavated for hidden information. Essentially, the quality of the output depends on the quality of input data. Alternately, a good analysis of faulty/bad data cannot result in meaningful outputs. The global challenge that arises during data analysis is the quality of data. Data quality is not intentionally reduced by unscrupulous systemic elements as inconsistencies have an uncanny way of creeping in due to various factors. The importance of data allows organizations to measure past performance quantitatively as well as to quantitatively ascertain present capabilities and thereby plan for future performance targets, leading to the study of data inconsistencies. This paper presents a conceptual outline of various categories and types of data inconsistencies, extending it further to briefly explain the data processing life cycle and the sources of data inconsistencies.

Keywords: Data Inconsistencies Sources, Data Processing, Data Science, Spatial, Temporal, Textual.

1. Introduction

With each passing second, every event that occurs in this world generates data. Different interpretations of the data are available, but David Hartzband describes the data as a collection of facts produced by direct or derived measurement or analysis that may be useful, redundant, and that must usually be interpreted in some method to be interpretable [9]. However, it is only recently, data recording and tabulation have gained importance.

Technological advancements in the field of Information Technology has made data capturing and recording a valued and one of the fastest emerging domains requiring experts and data scientists to decipher the information hidden within the raw data. Decreasing prices and increasing data capacities has to lead to data explosion [5]. This digital transformation of data has revolutionized not only the day to day activities but also business-related data. Traditional methods involve manually carrying out the tasks as compared to today's modern world, where machines and computers are dedicated, and specialized systems are

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Copyright © Authors

deployed. These modern-day machines monitor all the activities at a granular, and all related data is captured, thus generating humongous data. The continuous creation of data has posed new research challenges due to its complexity, diversity, and volume. Consequently, Big Data has increasingly become a fully recognized scientific field.

Various multi-scale applications on the cloud used in social networking like Facebook, YouTube, Google, Twitter, Whatsapp, etc., are becoming highly popular. They have changed the way that humans interact with each other and do business. These platforms are generating incredible and insane amounts of data requiring dedicated study, systems, and technological advancements in this field [8]. As per the most recent figures available which were released at the end of 2018, social media giants like Google have revealed that they process over 20petabytes of data per day [31], Facebook processes 2.5 billion pieces of a content exceeding 0.5 petabytes of data per day [25], YouTube uploads about one hour of video every second [34], Twitter, about 0.5 billion tweets per day [28] and in Astronomy, for example, satellites generate data in hundreds and thousands of petabytes. Every year there is a 40% rise in digital content. This includes not only an increase in internet users but also enterprises, organizations, education going online. To add to these vast amounts of data generated, IoT devices like sensors, surveillance cameras play a very crucial role [29]. It is estimated that by 2020, the digital universe would have generated 44 zettabytes of data [32].

Data generation has moved from linear to exponential, and now the exponential increase of data generated and collected has reached unprecedented and mind-boggling levels. This comes with its own challenges. It has been observed that data generated versus data recorded encompasses a delta error. Computer-based systems are termed as ‘unintelligent’ as they are incapable of deciphering ‘what is said and what it means’ in reality leading to data inconsistencies. The focus of the paper is to lay a conceptual baseline while defining the various taxonomies in data inconsistencies along with its subtypes in part 2,3 and 4 of this paper.

2. Data Analysis and Sources of Inconsistencies

To delve into the topic of data analysis, it is pertinent to know that data follows some basic rules in its raw and cryptic form and also that data can be broadly categorized into the following two groups:

- Domain Based: Globally, all business transactions are segmented between various types of businesses based on domains. Various studies have defined various domains [7, 10]. Different domains tend to generate similar data or data of similar nature. For example, data from the healthcare/medical domain have distinct features as compared to data from the human resources domain [20].
- Type Based: A second broad category of data is made based on its type [11]. Data could be in the form of numbers, text, audio, video, etc. [10].

2.1 Data Analysis and Processing

Many researchers have proposed and developed data anomaly detection algorithms and data processing tools based on the nomenclature of data. However diverse or alike these tools and techniques maybe they should rely on a high level of data processing steps as outlined below:

i. Data Collection: Basically, data processing starts with the collection of data. All transactions transpiring over the internet between two parties, entities or

Copyright © Authors

systems, leave a footprint of the transaction in the form of transaction type and the transfer of information between them. Next, the data collected in repositories in the form of databases, data farms, and business logic are programmed making the data virtual thereby increasing the quantity of data that can be collected and stored.

ii. Data Preparation: Though data collected in the previous stage is unintelligible there are robust systems that ensure complete capturing of source data. However, this captured data needs to be ‘sanitized’ before it can be subjected to various tools and techniques for further processing. This step is also called pre-processing of data. In this stage, there are two major activities firstly, data anomalies are detected and then the anomalies or data inconsistencies need to be eliminated so that only high quality and accurate data is pushed downstream for further processing.

iii. Data Input: Upon successful and ‘sufficient’ cleansing of data the stage can be considered set for the data to be further pushed for processing to extract/convert raw data into meaningful information. This cleansed data is warehoused to be retrievable/accessible by various tools for data processing or analysis. Data warehouses are complex storages, wherein data can be fed from various sources and used as a repository of combined data called big data [15].

iv. Data Processing: In this phase, the cleansed data is treated using various algorithms and systems to decipher the hidden information. Raw data subjected to different tools, techniques, and algorithms, results in information that has varied applications. Processing is performed utilizing machine learning techniques, but the method itself could differ slightly based on the original data being handled (data lakes, social networks, etc. and its expected usage (examining advertising patterns, the medical diagnosis from connected devices, determining customer needs, etc.). Popular open–source and closed–source systems/CRM systems like Sales force, Tableau, SPSS, Apache Spark, etc. are used for data processing [6].

v. Information Output: Processed data is called information. [Morrison, R. 2020] further mentions that converting or processing data results in information that can be further used for taking decisions or further downstream processing/consumption. It has been observed that the same data when processed using different tools gives rise to a different type of information. Essentially, this information is to be represented in a meaningful and usable manner for the intended purpose. Processed data can be represented in the form of graphs, videos, images, plain text, etc. Further, this information will lead to the requestor taking decisions based on a strong argument and not a speculative suggestion [15].

vi. Storage of Processed Data: The last phase of the data processing life cycle is the storage of the processes/analyzed data, which is nothing but insightful information useful for taking decisions. Storage of this information must be governed by rules which serve to ensure the protection of privacy and also to prevent misuse of the data by unscrupulous elements and hackers. This is governed by legal and statutory laws like General Data Protection Regulation (GDPR), Intellectual property rights (IPR), Information Technology (IT) Acts, etc. In addition to this, data must be stored in such a way that it can be indexed and categorized easily and efficiently so that it can be readily accessible to the right entity which demands it. The storage must also be cost-effective and must not be costlier than the data itself. If the data processing and storage is costlier in terms of monitory considerations as well as the importance, then it would lose its worth.

2.2 Sources of Inconsistencies in Big Data Process

Data collected into warehouses are from various sources. In big data, since several large datasets are combined to form a large hive of data stored in various geometric shapes and formats, data inconsistencies are bound to creep in.

The data preparation phase is the crucial step in big data processing as it deals with data sanitizing and cleaning, thus making the raw data collected into processable form. In this phase, data anomalies are identified and eliminated. These data anomalies or vulnerabilities are known as data inconsistencies.

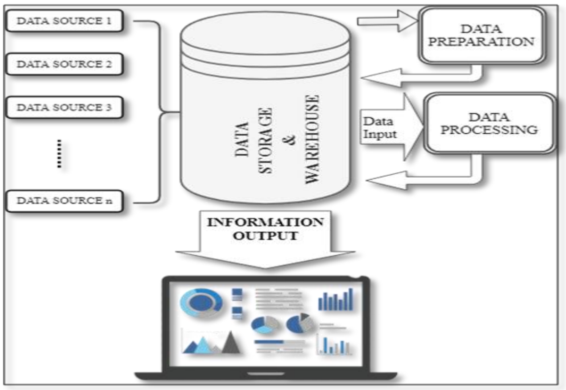

Figure 1. Model of Data processing System Architecture

Inconsistency present in data always creates issues in terms of data cleaning & integration as a part of big data analytics. Information and knowledge inconsistency is limitless and ambiguous in the actual globe and is known to be an acknowledged part of existence. In certain situations, the contradictory knowledge we find in our everyday lives is insignificant, triggering a slight inconvenience. The effect of such contradictory knowledge is significant and catastrophic, based on the type and the situations.

Data inconsistency is a state where multiple tables within a database deal with the same data received from different sources [8]. Also, data inconsistency occurs when multiple tables have similar data but in different formats. These definitions thus explain the main difference between data redundancy and data inconsistency [32].

Inconsistency is generally exacerbated by data redundancy. However, it differs from redundant data and its adjacent anomalies in that it usually applies to issues with the content of the database instead of its architecture and layout. This is where the presence of multiple channels and touch points, as well as the human ability to construct a creative spin on inputs, are beginning to create compound database difficulties.

3. Taxonomy of Data Inconsistencies

Before we dive deep into the discourse, it would be pertinent to understand the term inconsistency. The anomaly or infection of data can be in the form of data inconsistency, redundancy, and/or errors or defects.

- Errors / Defects: This would indicate that the data entered is incorrect by the user or that there is an error between the correct data and the data which has been entered or collected by the system [16].
- Redundancy: Data redundancy can be defined as the presence of the same data at multiple locations in the database [16].
- Inconsistency: Data inconsistency can be defined as the presence of data in dissimilar formats. It refers to a state of

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Copyright © Authors

availability or presence of the data in dissimilar formats. E.g. date can be represented in various formats, but a consistent format must be retained when recording data in a database. In case of any breach, it is termed as a data inconsistency. It leads to unreliable and meaningless information generated when the data is processed. [16].

Broadly data inconsistencies can be classified into four types namely: [5]

1. Temporal Inconsistencies: These types of inconsistencies relate to time. It contains conflicting information provided in big data. Despite conflicting data items embedded in large data with conflicting contexts, they can merge or overlap in time. Time internal conflicting data items may lead to partial temporal inconsistency or total temporal inconsistency. [25].
2. Spatial Inconsistencies: Spatial inconsistencies are those that generally occur in a data set which includes geometric or spatial dimensions. For spatial inconsistencies reduction, traditionally DB systems are getting enhanced to include spatially referenced data. Spatial inconsistencies can arise from:
    a. Geometric representation of objects
    b. The spatial relationship between objects
    c. Aggregation of composite objects
3. Text Inconsistencies: Text Inconsistencies are generally found in unstructured natural language text. Data generated from social media, blogs, emails, etc., are examples of text inconsistencies [28]. In big data, if two texts relate to the very same activity or event, they were said to have been co-referenced. [34]. There are multiple ways to detect text inconsistencies, but contradiction detection is one of the well–known methods that detect text inconsistencies having many applications [34].
4. Functional Dependency Inconsistencies: Related values to parameters with correlation must also show similar correlations. Such dependency where one attribute depends on others is known as Functional Dependency inconsistency [29]. Since we have shown, vast datasets are maintained, aggregated and cleaned with the aid of the Relational Database Management System (RDBMS), whereby functional dependencies performs a key role in maintaining the integrity of the datasets to achieve functional dependence. [34]

3.1 Types of Temporal Inconsistencies

Temporal inconsistencies are classified into the following subtypes:

Consider Event A→ Ask a user to fill out a form

Event B→ Save the form which the user fills out

i. Partial: In this category of inconsistency, the time ranges of two inconsistent occurrences overlap partially. E.g. If interval A overlaps with interval B, then add both A and B to the resulting set of intervals that overlap.

ii. Complete: In this category of inconsistency, time intervals of two inconsistent occurrences coincide, containment. If (StartDateA<= EndDateB) and (EndDateA>= StartDateB) is false then inconsistency occurs.

iii. Anomalous Value: In this type of inconsistency time-series data has an anomalous value and stands out of the sequence. E.g. 2, 4, 6, 8, 10, 23000, 12, 14, 16… Thus here 23000 is the inconsistency.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Copyright © Authors

iv. Contextual: In this category of inconsistency, time-series data has an artifact in a specified sense. For example, if the correct context cluster can be identified and if the instance is far from any of the current metric patterns for that cluster, that instance can be classified as an anomaly. An instance is therefore anomalous if, provided its context, its metrics indicate unusual activity.

v. Motif: In this category of time series inconsistency data, there is a section of data which recurs and is anomalous. E.g. revenue at a store every day is a time-series data at a day level. Time series is any data that is associated with time (daily, hourly, monthly, etc.)

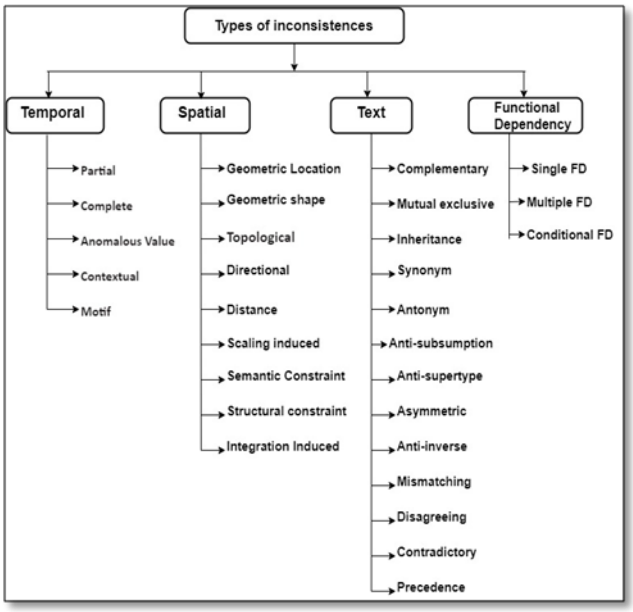

Figure 2. Type of Inconsistencies

3.2 Types of Spatial Inconsistencies

Under the types of spatial inconsistencies, these are the following subtypes defined:

i. Geometric Location: In this type of inconsistency, a spatial object has a conflicting geometric location, e.g. a live traffic system/road map.

ii. Geometric shape: In this type of inconsistency a spatial object has a conflicting geometrical shape, e.g. the shape of a landmass as captured by a satellite.

iii. Topological: In this type of inconsistency, there is a violation of the topological constraint. E.g. a lot of errors that can be found in GIS data sets arise due to lack of knowledge about topological relations between the geographical objects stored in the database.

iv. Directional: In this type of inconsistency, there is conflict in the reporting of directional correlation between two objects. E.g. at one instance, A is directed towards B, whereas at another location it is represented as B is directed towards A. While it may not always be a direct opposite, it could be in violation by a few degrees also and would still qualify as directional inconsistency.

v. Distance: In this type of inconsistency, there is a conflict in the reportedly stored/represented data regarding distances between two points in space which may be collected by the geographic information system (GIS), e.g. distance points A and B is reported as different by a few meters or more and stored at different locations in the database.

vi. Scaling Induced: In this type of inconsistency, there is an inconsistency observed in the finding when the scales/systems/tools used are different. E.g. an algorithm A is replaced by another algorithm B to carry out a similar operation, thereby resulting in inconsistency in computed results.

vii. Semantic Constraint: In this type of inconsistency, there is an inconsistency due to the reference of an object by different synonymous

Copyright © Authors

names, e.g. a child can be called a kid, a boy, etc.

viii. Structural Constraint: In this type of inconsistency there is a violation of structural integrity constraint of geometric primitives [2], e.g. the set of all employees work for organization A. However organization A does not work for B if one employee also works for organization B.

ix. Integration Induced: In this type of inconsistency, different representation of the same spatial object from different sources resulting in violation of the constraint that object must have unique geometric representation, e.g. two systems A and B collecting the email address and mobile number and storing in the data warehouse as conflicting mobile numbers or email addresses for the same person.

3.3 Types of Text Inconsistencies

Under the types of Text-based inconsistencies, the following subtypes are defined:

i. Complementary: In this type of inconsistency, it is observed that data presented is complementary at different locations, e.g. A is related to B, and A is not related to B.

ii. Mutual exclusive: In this type of inconsistency, the same entity A is reported as two mutually exclusive entities, e.g. A is a male, and A is a female.

iii. Inheritance: In this type of inconsistency, there is a conflict between the definition and the specific instance depicted, e.g. claim made is a penguin cannot fly, however, a penguin is a bird and birds can fly, thus penguins can fly.

iv. Synonym: In this type of “inconsistency” multiple instances of an entity are present but the terms associated are synonyms of each other. While some may not consider it an inconsistency as the base meaning remains unchanged, for all practical purposes this is an inconsistency. For e.g. P5 is a very fast processor. The processing speed of P5 is very rapid.

v. Antonym: In this type of inconsistency there is serious representation as the meaning of the statement is completely reversed due to the use of opposite words. E.g. A is a boy. A is not a boy.

vi. Anti – subsumption: In this type of inconsistency we observe that while there is a correlation or subsumption between two words the references made are not in agreement. For e.g. A is a surgeon but A is not a doctor.

vii. Anti – supertype: In this type of inconsistency there is a mismatch between the definition of the same elements while defining its category e.g. A is defined as a robot and also a four-legged support system

viii. Asymmetric: In this type of inconsistency there is a bidirectional anomaly between the references made e.g. A is related to B but B is not related to A.

ix. Anti – Inverse: In this type of inconsistency which is very close to the anti – inverse type, there is a disagreement between relationship operator between two elements of related data sets e.g. A contains B but B is not contained in A as in Ram is Hari’s father but Hari is not Ram’s child.

x. Mismatching: In this type of inconsistency there is an outright mismatch or disagreement when referring to the elements e.g. application A works on a platform B but platform B does not support application A.

xi. Disagreeing: In this type of inconsistency two conflicting responses are recorded while both may be correct e.g. A has a capacity of N GB and in another reference, A has a capacity of M GB where M ≠ N

xii. Contradictory: In this type of inconsistency there is a factual anomaly reported/recorded e.g. A was born on Date1 and A was born on Date2 where Date1 ≠ Date2.

xiii. Precedence: In this type of inconsistency established past precedence is violated by way of the following statement e.g. A is the father of B but B was born before A. Another example can be A became the 1st president and B become the 2nd president has an anomalous statement downstream which says that President 2 was succeeded by President 1.

3.4 Types of Functional Dependency Inconsistencies

Under the types of Functional Dependencies in relational databases the inconsistencies observed are of the following subtypes:

1. Simple FD: In this type of inconsistency if R is a relation with attributes X and Y, a functional dependency between the attributes is represented as X→Y, which specifies Y is functionally dependent on X. Here X is a determinant set and, Y is a dependent attribute. Each value of X is associated with precisely one Y value. In this context, if there is a violation in the equation X→Y, then it is termed as a single FD type of inconsistency, e.g. there are multiple responses observed for the same determinant variable which would factually not be plausible. [33]

2. Multiple FD: This type of inconsistency is also called multi-valued dependency. In this type of dependency, there is more than one dependent variable on the determinant variable. In simple words, any change in the determinant variable, there is a corresponding change observed in the responses of more than one variable. We can represent an equation as Y = f(X), where f(X) = f(x1, x2, x3… xn). The anomaly is observed when there is a change in Y, which is not impactful in any given response variable. For e.g. a course can have multiple books written by different authors. If a new book is added to the course, then a new title and a corresponding author must be added to the record. If a new course were to be added and only the title was appended but the author was not appended or vice – versa then it would constitute a multiple functional dependency type of inconsistency.

3. Conditional FD: This type of inconsistency is an extension of the multiple functional dependency type of inconsistency. In this anomaly, there is a linking or conditional relation that exists between the response variables. If X is a determinant variable that has A and B as a response variable and B is dependent on the constraint laid out based on A, then this is called conditional functional dependency. E.g. A can either be fruits or vegetables, and then B will be dependent on the value taken by A. If A is a fruit, then B cannot have a vegetable as a value. Another example can be the dependency of the phone number and country code. Depending on the country code, the phone numbers can only take a certain set of valid values. Any anomaly with regards to this is called a conditional

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Copyright © Authors

functional dependency type of inconsistency [33].

4. Conclusion

Having researched various aspects of inconsistencies, we can conclude that data inconsistencies are importantly not random; rather they can be classified under different taxonomical groups. An attempt has been made to mention in brief about the data processing life cycle, which is a universal framework that can be further modified based on the domain, technology, or type of task at hand. A brief mention of the sources of data inconsistencies has also been made. In this paper, data inconsistencies have been elaborately listed out and categorized. Further, with the help of relevant examples, the concepts regarding the type of data inconsistencies have been detailed. This paper indeed sets a foundation that can be a relevant reference paper for all researchers desirous of creating tools, techniques, systems, and algorithms for anomaly detection and removal of data inconsistencies in Big Data.

References

[1] Bohannon, P., Fan W., Geerts F., Jia X., Kementsietsidis A., "Conditional Functional Dependencies for Data Cleaning," University of Edinburg research publications.

[2] Bresina, J L, Morris P H, (2006), "Explanations and Recommendations for Temporal Inconsistencies", IWPSS, https://www.stsci.edu/largefiles/iwpss/20066061912IWPSS_draft4.pdf

[3] Brisaboa, Nieves and Luaces, Miguel and Rodriguez, Andrea and Seco, Diego. (2014). "An inconsistency measure of spatial data sets with eespect to topological constraints," International Journal of Geographical Science. 28. 56-82. 10.1080/13658816.2013.811243.

[4] Dr. S. Vijayarani and Ms. S. Sharmila, "RESEARCH IN BIG DATA: AN OVERVIEW," Informatics Engineering, an International Journal (IEIJ), Vol.4, No.3, September 2016

[5] Du Zhang, 'Inconsistencies in Big Data' proceeding, Cognitive Informatics & Cognitive Computing (ICCI*CC), 2013 P. 61-67 12th IEEE Conference.

[6] Garboden, Philip. (2020). "Sources and Types of Big Data for Macroeconomic Forecasting," 10.1007/978-3-030-31150-6_1.

[7] Hartzband, David. (2019). "What Is Data?" DOI: 10.4324/9780429061219-2.

[8] Jeffrey Ray, Olayinka Johnny, Marcello Trovati, Stelios Sotiriadis and Nik Bessis, "The Rise of Big Data Science: A Survey of Techniques, Methods and Approaches in the Field of Natural Language Processing and Network Theory," Big Data Cogn. Comput. 2018, 2, 22; doi:10.3390/bdcc2030022.

[9] Khan, Samiya and Liu, Xiufeng and Shakil, Kashish and Alam, Mansaf. (2017). "A survey on scholarly data: From big data perspective," Information Processing and Management.DOI53. 923-944. 10.1016/j.ipm.2017.03.006.

[10] Krogh, Jesper. (2020). "Data Types," DOI 10.1007/978-1-4842-5584-1_13.

[11] Kumar, Praveen. (2019). "BIG DATA ANALYTICS IN HR DOMAIN", DOI 10.1729/Journal.22887.

[12] M. V. Martinez, A. Pugliese, G. I. Simari, V. S. Subrahmanian, and H. Prade, "How dirty is your relational database? An axiomatic approach," in Proc. 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ammamet, Tunisia, LNAI 4724, 2007, pp.103-114.

Copyright © Authors

[13] M-C de Marneffe, A. N. Rafferty and C. D. Manning, Finding Contradictions in Text, Proc. of 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2008, pp.1039-1047.

[14] Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali,1 Muhammad Alam,Muhammad Shiraz,1 and Abdullah Gani, “Big Data: Survey, Technologies, Opportunities, and Challenges,” Hindawi Publishing Corporation, The Scientific World Journal, Volume 2014, Article ID 712826,https://doi.org/10.1155/2014/712826

[15] Özsu, M. and Valduriez, Patrick. (2020). “Big Data Processing.” 10.1007/978-3-030-26253-2_10.

[16] Ptiček, Marina and Vrdoljak, Boris. (2018). “Semantic web technologies and big data warehousing”. 1214-1219. 10.23919/MIPRO.2018.8400220.

[17] Ritter, D. Downey, S. Soderland and O. Etzioni, It’s a Contradiction-No, It’s Not: A Case Study Using Functional Relations, Proc. of Conference on Empirical Methods in Natural Language Processing, 2008.

[18] Samiddha Mukherjee, Ravi Shaw, “Big Data – Concepts, Applications, Challenge and Future Scope,” International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016, ISSN (Online) 2278 – 1021, ISSN (Print) 2319 – 5940

[19] Sergio Luján-Mora, Manuel Palomar, ‘Reducing Inconsistency in Integrating Data from Different Sources’. Proceedings 2001 International Database Engineering and Applications Symposium, p. 209-218: IEEE Computer Society, Grenoble, July 16-18 2001. https://doi.org/10.1109/IDEAS.2001.938087

[20] Smirnov, Alexander &Levashova, Tatiana &Shilov, Nikolay. (2012). Ontology Alignment for IT Integration in Business Domains. 127. 153-164. 10.1007/978-3-642-34228-8_15.

[21] Yaqoob, Ibrar& Hashem, Ibrahim and Gani, Abdullah & Mokhtar, Salimah& Ahmed, Ejaz and Anuar, Nor and Vasilakos, Athanasios. (2016). “Big Data: From Beginning to Future.” International Journal of Information Management. 36. 10.1016/j.ijinfomgt.2016.07.009.

[22] Zhang, “On Temporal Properties of Knowledge Base Inconsistency.” Springer Transactions on Computational Science V, LNCS 5540, 2009, pp.20-37.

[23] https://pediaa.com/what-is-the-difference-between-data-redundancy-and-data-inconsistency Accessed on 2/3/2020.

[24] https://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/ Accessed on 2/3/2020.

[25] https://www.bizdata.com.au/blogpost.php?p=costs-of-data-redundancy-and-data-inconsistency Accessed on 2/3/2020.

[26] http://www.brainkart.com/article/Types-of-Relationships-in-a-database_37285/ Read on 26/01/2020.

[27] https://www.dsayce.com/social-media/tweets-day/26/01/2020.

[28] https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm26/01/2020.

[29] https://www.happiestminds.com/Insights/big-data-analytics/ Accessed on 2/3/2020.

Copyright © Authors

[30] https://www.heshmore.com/how-much-data-does-google-handle/ Accessed on 2/3/2020.

[31] https://www.kdnuggets.com/2012/12/idc-digital-universe-2020.html Read on 26/01/2020.

[32] https://www.techopedia.com/definition/19504/functional-dependency Read on 26/01/2020.

[33] https://www.washingtonpost.com/national/health-science/ Accessed on 2/3/2020.

Copyright © Authors

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731