

Using Web Meta-data to Discover Patterns of User Behaviour

Weiam S. Elsaghair^{1*}

¹ Kadir Has University, * Corresponding author, welsaghair@gmail.com

Abstract

Getting specific and accurate data through discovering patterns of usage mining depends on a good description of the used data will be provided in this paper. Also, web metadata will describe the data which is used in pattern discovery. So, using web data to discover user patterns by analyzing user behaviour is not only useful to infer and interpret the expected information about predictive patterns. But also offers a great deal by improvement of services, data flow and web warehousing in the context of big data.

Keywords: Web mining, Metadata, Big data, User patterns and behaviour, Web analytics.

Citation: Elsaghair, Weiam S. (2018, October) *Using Web Meta-data to Discover Patterns of User Behaviour*. Paper presented at the Fifth International Management Information Systems Conference.

Editor: H. Kemal İlter, Ankara Yıldırım Beyazıt University, Turkey

Received: August 19, 2018, **Accepted:** October 18, 2018, **Published:** November 10, 2018

Copyright: © 2018 IMISC Elsaghair. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Using Web Meta data to discover patterns of user behaviour

Weiam S. Elsaghair\ Kadir Has University-Istanbul

The graduate school of science and engineering

Management Information Systems. Department-2018

Abstract: Getting specific and accurate data through discovering patterns of usage mining depends on a good description of the used data will be provided in this paper. Also, web metadata will describe the data which is used in pattern discovery. So, using web data to discover user patterns by analyzing user behaviour is not only useful to infer and interpret the expected information about predictive patterns. But also offers a great deal by improvement of services, data flow and web warehousing in the context of big data.

Keywords: Web mining, Metadata, Big data, User patterns and behaviour, Web analytics.

Introduction:

The Internet popularity is increasing and the rapid development of E-commerce, also the Internet-based businesses' websites are facing increasing competition. E-commerce sites generate large amounts of data. These data including potential consumer-related information is valuable for market analysis and prediction. In last times new concepts have appeared. Those concepts concerned about how to manage and analyze data and get beneficial results. Data science defined as how to get and understand the data and extract their value.

Therefore, getting the desired data which is provided by business analytics. This data is important for data science, so it becomes important in the last 10 years. Business analytics is defined as a result of using tools and techniques to turn data into meaningful business insights.

Business analytics is used to describe all functions by developing skills, techniques and data mining algorithms then generate valuable information which it will be in a represented and high readable format for managers to take control and manage business operations. When Business analytics functions work effectively and efficiently will become a core competency for the organization in the form of valuable business intelligence that will support the strategic actions undertaken by the organization.

Data mining considered as an essential factor which is lead the applications of business analytics. It uses to analyze data to get a better understanding and prediction of future outcomes. Therefore web mining techniques are used to extract beneficial information. This paper aimed to focus on using metadata to describe used data in analysis techniques and clarify their advantages which make a data

more clear and effective also create associations and relationships between items. According to the previous, we can categorize web mining techniques into three types:

- Web content mining.
- Web structure mining.
- Web usage mining.

Those categories based on a mining operation, whereas web content mining is used to describe that extracted information from web content. Web content consists of several types of data such as textual, semi-structured data and HTML documents. Also, web structure mining uses hyperlinks to catch meaningful information, whereas it is possible to arrange the websites and deduct the relationship between links. Furthermore, web usage mining is beneficial to gain meaningful patterns from web data, these data it could be web server access logs, user sessions or transactions, proxy server logs, browser logs, user profile, registration data, cookies, user queries, bookmarks data, mouse clicks and scrolls, and any other data as a result of interactions.

Problem definition and Scope:

This paper will address a web usage mining and extract patterns which are beneficial to analyze user's behaviour:

- Discover and analyze patterns of customer's behaviours according to their browsing and session time, then Determine a predicated customers.
- The most hit Links.
- The most active users.
- Develop log server metadata framework.
- In this paper, we will focus on web usage mining, whereas a customer business activities and browsing data are preserved in logs files. And then analyze these data and use it to understand and handle customer's web behaviour.

Metadata: Simply is a data which is used to describe used data. That helps to organize and a better understanding of data. For instance, information about a text document's metadata may contain the title, subject, author..etc. An image which is may include metadata that describes the size of the image, the colour, resolution, and creation date. So, the typical metadata elements may have the following elements:

- Title and description.
- Tags and categories.

- Who Created and when.
- Last modification and when.
- Who can access and update?

Metadata is not a new concept it used in libraries for classifying and categorize books based on title, author and subject so it is easy to do with your searching. Now, by describing HTML elements that directly communicate and clarify website information, can improve the effective search engine optimization, also create associations and relationships between items. The value of metadata could be the ability to more efficiently classifying and organize information, more intelligence and higher quality information to support big data initiatives, automation, compliance, data sharing and others. Also, Customer behaviour, navigation through websites and visitors activity can be measured and analyzed objectively. The massive intention describes the customer activities inside the website and the potential customer that is through using the qualitative method to analyse web server logs, web client logs, proxy server logs, and cookies data. And using some mining techniques as follows:

1-Association Rules:

The most popular method to discover the interesting relations between variables in the large database based on the concept of strong rules is an association rule.

2-classification and Prediction:

A Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data.

Metadata Framework Proposal:

To better understanding of a concept of metadata. So it is helpful to categorise it into five sections as the following table:

Type	Description
Descriptive	It used to identify and describe the object. For example the title, subject, date, description, type of resource.
Structure	It is about how to Describe the descriptive metadata and determine what that attribute means. (chapter number, indexes, pages, sections, tables of contents)
Administrative	Provides information to help resource managing such as (when and who it was created, file type and other technical information.
Preservation	Is information that supports and documents the digital preservation process, and stores technical details on the format structure and use of the digital content (backup, refreshing cycles, migration..etc).
Technical	It is related to systems behaviour from business and functional specifications. (Accessibility, security issues, hardware and software documentation).

Table (1) Metadata types.

This paper adapts web server logs and client-server logs to achieve the following data elements:

IP address: it is a text data type which is used to determine user surfing percentage according to their regions. This element will be beneficial in association and sequential rules.

URL Browsing: Also By using association and sequential rules to analyze browsing url we can catch a user orientation and traffic pages to improve the site works.

Timestamp: it is a date time data type and gives information about the duration of the user session.

Session ID: Integer data type as a unique element for a unique session.

Browsing Bytes: it is about the size of transmitted bytes. And it is a double data type.

User agent string: Describing a series of data by text data type this series has a user browser type and user operating system. It will be classified by using a decision tree algorithm it will be usable method in this paper.

Logs data metadata framework:

	Structure	IP address (Internet Protocol User unique address)	URL (Uniform resource locator)	TimeStamp (Current time of an event recorded by computer)	Session ID (Unique identifier for a specific user)	Browsing Bytes (Transmitted bytes)	User agent string (Technical information)
Types of metadata	Administrative	Url Path. File source (Digital media, Printed stuff). File types: Text, video, photo, Audio..etc File location. File size. File description (Image resolution, dimensions ...etc). When and who created. Last modification date. Access authority.					
	Preservation	Provenance. Authenticity. Preservation Activity. Technical Environment. Rights Management. Backup Plan, Preservation events.					
	Technical	Database Owner. Physical characteristics (table space, extents, segments, blocks, partitions). Performance tuning (processors, indexing). Tables specifications and columns details. Constraints, the relationship between tables. Data dictionary. Software specifications, versions and documentation. Compatibility. Authentications and security issues e.g. formats, encryption keys, passwords. Tracking of system response time.					
	Descriptive	String. Not Null Indexing	String. Not Null Indexing	Date Time. Not Null Indexing	Integer. Not Null. Indexing. <ul style="list-style-type: none"> • Unique. • Auto Increment 	Float. Not Null Indexing	String. Not Null Indexing

Table (2) Metadata types with logs data.

characteristicsAttributesmetadata and	Source of metadata.	<ul style="list-style-type: none"> • Internal metadata Generated by the creator. (File name, File format) • External metadata Created later by the someone else original creator (URLs, Rights and other legal information.
	Method of metadata creation.	<ul style="list-style-type: none"> • Automatic Metadata generated by a computer. (User transaction logs). • Manual metadata created by humans. (Descriptive metadata).
	Nature of metadata.	<ul style="list-style-type: none"> • Non-expert metadata Created by persons. • Expert metadata Created by experts.
	Status.	<ul style="list-style-type: none"> • Static. • Dynamic. • Long-term metadata. • Short-term metadata.
	Structure.	<ul style="list-style-type: none"> • Structured metadata (Predict standards). • Unstructured metadata (note fields, and other free text annotations).
	Semantics.	<ul style="list-style-type: none"> • Controlled metadata (standard vocabulary and authority). • Uncontrolled metadata.

Table (3) Metadata attributes and characteristics.

Use of Relevant Web Mining Techniques from a Metadata Perspective:

1. Association Rules

The most popular method to discover the interesting relations between variables in the large database based on the concept of strong rules is an association rule.

In following the problem of association rule mining is defined as:

Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of binary attributes called items.

Let $D = \{t_1, t_2, \dots, t_m\}$ be a set of transaction called the database.

Each transaction in has a unique transaction ID and contains a subset of the items in a rule is defined as an implication of the form.

$$X \Rightarrow Y$$

Where $X, Y \subset I$ and $X \cap Y = \emptyset$

Every rule is composed of two different sets of items, also known as item sets, X and Y, where X is called antecedent or left-hand-side (LHS) and consequent or right-hand-side (RHS).

Interesting Concepts:

Support: How frequently a specific items set occurs in the database (the percentage of transactions that conation all of the items in the item set).

So, The Support value of $X = \text{sup}(X)/N$.

Which X contains the item-set, N the transactions.

For Example The item-set $\{I1, I2\}$ has support of $1/5 = 0.2$, It occurs in 20% of all transactions (1 out of 5 transactions).

Confidence: The probability that items in RHS will occur given that the items in LHS occur. It is computed as $\text{Support (LHS} \cup \text{RHS)} / \text{support (LHS)}$.

For Example: The rule $\{a, b\} \Rightarrow c$. (which a, b, c is items for sale) has a confidence of $0.2/0.2 = 1$ in the database, which means that for 100% of the transactions containing a and b the rule is correct (100% of the times a customer buys a and b , c is bought also).

2. classification and Prediction:

A Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data.

Prediction is one of the most important usages of decision tree models. Using the tree model derived from historical data, it's easy to predict the result for future records.

A Decision tree is a common data mining method for establishing classification systems based on multiple covariates or for developing prediction algorithms for a target variable.

$$(\) = - \sum_{i=1}^n 2(\)$$

Pattern Discovery:

Association Rules: in association rules, the dataset was prepared from server logs file, which has about 14600 records. It is used to analyse data and extracts the associations.

10.130.2.1,[30/Nov/2017:14:26:05,GET /css/normalize.css HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:26:05,GET /css/main.css HTTP/1.1,200
10.131.2.1,[30/Nov/2017:14:26:05,GET /css/style.css HTTP/1.1,200
10.131.0.1,[30/Nov/2017:14:26:05,GET /js/vendor/modernizr-2.8.3.min.js HTTP/1.1,200
10.130.2.1,[30/Nov/2017:14:26:06,GET /js/vendor/jquery-1.12.0.min.js HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:26:06,GET /bootstrap-3.3.7/js/bootstrap.min.js HTTP/1.1,200
10.130.2.1,[30/Nov/2017:14:26:07,GET /fonts/fontawesome-webfont.woff2?v=4.6.3 HTTP/1.1,200
10.130.2.1,[30/Nov/2017:14:26:12,GET /sign.php HTTP/1.1,200
10.131.0.1,[30/Nov/2017:14:26:49,POST /action.php HTTP/1.1,302
10.131.0.1,[30/Nov/2017:14:26:51,GET /login.php HTTP/1.1,200
10.131.0.1,[30/Nov/2017:14:27:00,POST /process.php HTTP/1.1,302
10.131.0.1,[30/Nov/2017:14:27:00,GET /home.php HTTP/1.1,200
10.131.0.1,[30/Nov/2017:14:27:01,GET /bootstrap-3.3.7/js/bootstrap.js HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:27:01,GET /js/vendor/moment.min.js HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:27:15,GET /contest.php HTTP/1.1,200
10.131.2.1,[30/Nov/2017:14:27:44,GET /contestproblem.php?name=RUET%20OJ%20Server%20Testing%20Contest HTTP/1.1,302
10.130.2.1,[30/Nov/2017:14:27:45,GET /countdown.php?name=RUET%20OJ%20Server%20Testing%20Contest HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:27:52,GET / HTTP/1.1,302
10.128.2.1,[30/Nov/2017:14:27:53,GET /login.php HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:27:53,GET /css/bootstrap.min.css HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:27:53,GET /css/font-awesome.min.css HTTP/1.1,200
10.131.2.1,[30/Nov/2017:14:27:53,GET /css/normalize.css HTTP/1.1,200
10.129.2.1,[30/Nov/2017:14:27:53,GET /css/style.css HTTP/1.1,200

Table (4) Sample of server logs

To prepare association dataset it is important to add transaction-id which is a numerical count of the transactions, and transaction-sequence which describes the sequence of URLs.

1	URL1	GET /css/normalize.css HTTP/1.1
2	URL2	GET /css/bootstrap.min.css HTTP/1.1
3	URL3	GET /css/font-awesome.min.css HTTP/1.1
4	URL4	GET /css/main.css HTTP/1.1
5	URL5	GET /css/style.css HTTP/1.1

Table (5) describe the transaction sequence attribute.

Table (6) describes the discovered rules. The association between Url4 and Url1, Url3 is strong which means if the user goes to the Url1 100% of them also visit Url2, Url3 links. Such rules are of 0.7% among all the web traffic transactions. Another rule addresses that 87% of users visit Url5 before they visit Url4 site, such as rules are 23% among all the web.

Rule	Confidence	Support
Url3->Url1&Url2	75%	11%
Url4->Url1&Url2	%111	%1.0
Url2->Url1&Url5	%03	%22
Url5->Url1&Url1	%03	%22
Url4->Url5&Url2	%20	%23
Url1&Url3->Url2	%01	%1.0
Url1&Url4->Url2	%25	%11
Url1&Url5->Url2	%111	%1.0
Url2&Url3->Url1	%33	%22
Url2&Url4->Url1	%25	%11
Url2&Url5->Url1	%111	%1.0
Url5&Url3->Url1	%33	%22
Url5&Url4->Url1	%05	%1.5
Url5&Url3->Url2	%111	%1.4
Url5&Url4->Url2	%05	%1.5
Url1&Url2->Url5	%111	%1.4
Url1&Url4->Url5	%00	%22
Url2&Url4->Url5	%01	%1.0

Table (6) Discovered rules.

Classification: To classify the user's according to according to a decision tree it is important to rely on a base for this classification. For instance

Categories the user's according to their gender or age and identify them according to specific criteria like finishing the buying procedure. That all can help to make a decision like 50% of male in middle age they will buy or 70% of female in middle age they will buy. Since this data is not available in the dataset so it is impossible to find this kind of classification.

The most hit links it could be represented to determine which site has the more visiting amount.

Page	NumberOfhits
1.s TTH h7E?i?php.tG bss/ GEG	1
aTT HTT7T7sc?E h.EaTT bss/ GEG 1.s	414
1.s ?sT hlpE?i?pTTps.Hclc bss/ GEG	1
c.pTT pE h.EaTT bss/ GEG-1.s aTT T.7	333
1.s csaihhpE?i? bss/ GEG	232
GET /img/ruet.png HTTP/1.1	213
1.s aT.7pT7E?i? bss/ GEG	242
1.s ?sT hlpE?i?pTTps.?TsTh.li bss/ GEG	3
1.s h.pppE?i? bss/ GEG	4
GET /css/normalize.css HTTP/1.1	412
1.s pp7chITE?i?p.c p.?ipTTHTcspeaTp.t bss/ GEG	3
1.s uT usTpsiE h.EuT bss/ GEG	50
1.s ITITT7E?i? bss/ GEG	2
1.s uT aics7E h.EuT bss/ GEG	52
bss/ GEG uT HTT7T7sc?EuT 3.3.0-1.s HTT7T7sc?	131
GET /standings.php?id=16 HTTP/1.1	154
1.s aTT ch.EaTT bss/ GEG	334
/.1s T?pc7pE?i? bss/ GEG	0

GET /robots.txt HTTP/1.1	224
1.s Thl.E?i? bss/ GEG	124
GET /css/style.css HTTP/1.1	335
b.1/ IT pE?i? bss/ GEG	4

Table (7) Pages hitting rates

Also, most active users it could represent to determine user's activation during links.

User IP	NumberOfhits
11.122.2.1	4250
11.131.2.1	4150
11.131.2.1	1020
11.123.2.1	1052
11.131.1.1	4132

Table (8) the most active users

As a result by describing a dataset using Metadata as mentioned before in table 2, 3, will give an advantage to a better understanding of a dataset and use it in a big data analytics and open data with supporting many concepts as following:

- **Consistency:** combines data from different resources.
- **Availability:** a description of the dataset that is included in the search.
- **Competences:** a description of data that is generated in real time that includes the date and time of last modification.
- **Conformance:** a description of a dataset according to explicit rules or standards.
- **Credibility:** a description of trustworthy sources.
- **Processability:** a description to which it can be understood and handled by automated processes.
- **Relevance:** contains the necessary information to support the application.
- **Interoperability:** the ability of a system or a product to work with other systems or products without special effort on the part of the customer.
- **Transparency:** the data stream being sent and stream being delivered are same, and consumers need to be aware of what is collected and shared to ensure checks and balances and better exchange of value with data collectors.
- **Customer data-driven:** making strategic decisions based on data analysis and interpretation with access when they need it.

Conclusion:

The web is the most important techniques in business and commerce. Therefore designing and managing the websites have a great impact to increase their visitors and improve business functions. So this paper conducts a vision of what metadata is and how we can use it to improve the clarity and understanding of data. So by describing different types of the dataset (structured, semi-structured and unstructured) by Metadata, it is possible to extract a meaningful knowledge from big data. Metadata provides the value and purpose of the data content, and thus becomes an effective tool for quickly locating information a must for big data analytics and business user reporting.

References:

11. Tawfiq, A. Al-said, Ahmed J. Obaid. Discovering similar user navigation behaviour in Web log data(2016). International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 16. Retrieved from <https://www.researchgate.net/publication/308114719>.
2. Jaideep Srivastava , Robert Cooley , Mukund Deshpande, Pang-Ning Tan. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data(2000), University of Minnesot Volume 1, Issue 2. Retrieved from <http://yildiz.edu.tr/~aktas/courses/CE-0114890/g10-p3.pdf> .
3. P. Jomsri. Improving the Performance of Proxy Server by Using Data Mining Technique(2013). International Journal of Computer and Information Engineering Vol:7, No:7. Retrieved from <https://waset.org/publications/16471>.
4. R. Cooley, B. Mobasher, and J. Srivastava. Web Mining: Information and Pattern Discovery on the World Wide Web (1997). , the University of Minnesota Minneapolis, MN 55455, USA. Retrieved from <https://pdfs.semanticscholar.org/6a13/59439d4a745a31e1eca2d55688132f7e0fb7.pdf>
5. Pui Mun Lee, SIM University [UniSIM], Singapore. Use Of Data Mining In Business Analytics To Support Business Competitiveness. Review of Business Information Systems (2013). Second Quarter Volume 17, Number 2. Retrieved from <http://clutejournals.com/index.php/RBIS/article/download>
6. Mariusz Łapczyński and Sylwester Białowas. Discovering Patterns Of User's Behaviour In An E-shop-comparison of consumer buying behaviours in Poland and other European countries(2013). Cracow University of Economics Poland. DOI: 10.1.1.872.9314.
7. Gan Teck Wei, Shirley Kho, Wahidah Husain, Zurinahni Zainol. A Study of Customer Behaviour Through Web Mining. Journal of Information Sciences and Computing Technologies(2015) .DOI: 10.1.1.675.4784
8. Xiaolong Zhang, Wenjuan Gong, Yoshihiro Kawamura. Customer Behavior Pattern Discovering With Web Mining (2004). Wuhan University Of Science and Technology. Retrieved from <https://pdfs.semanticscholar.org/5909/ef17a6308276e051f517b38c026d3504d47d.pdf>.
9. V. Sathiyamoorthi, Dr.Murali Bhaskaran. Data Preprocessing Techniques For Pre-Fetching And Caching of Web Data Through Proxy Server November (2011). IJCSNS International Journal of Computer Science and Network Security.VOL.11 No.11. November. Retrieved from <https://pdfs.semanticscholar.org/8a78/e9dad3fae79843fbc7b988565822a21e2d79.pdf>.
10. John Cleary, Geoffrey Holmes, Sally Jo Cunningham, Ian H. Witten. MetaData For Database Mining. University of Waikato Hamilton, New Zealand. Retrieved from <https://pdfs.semanticscholar.org/768a/e2ea1cf47b2ac4b8e3059f639586541b4f48.pdf>
11. Jessica Venlet, Karen Stoll Farrell, Tammy Kim, Allison Jai O'Dell, Jackie Dooley. Descriptive Metadata For Web Archiving Literature Review Of User Needs(2018). OCLC Research Dublin, Ohio 43017 USA. DOI: 10.25333/C33P7Z.

12. RLG. Descriptive Metadata Guidelines For RLG Cultural Materials. Mountain View (2005). California jan . Retrieved from https://www.oclc.org/content/dam/research/activities/culturalmaterials/RLG_desc_metadata.pdf.
13. John Stamey, Jean-Louis Lassez, Daniel Boorn, Ryan Rossi. Client-Side Dynamic Metadata In Web 2.0. Coastal Carolina University. Retrieved from http://ryanrossi.com/pubs/SIGDOC2007_Dynamic_Metadata.pdf.
14. Atalm Pre-Conference. Types Of Metadata(2017). Image Digitization on a Budget. Retrieved from <http://www.sustainableheritagenetwork.org/digital-heritage/2017-pre-atalm-shn-workshop-image-digitization-budget-slides>.
15. Mr.akshay, mr.Balram Pmswani. Web Usage Mining Has Patterns Discovery(2013). International Journal Of Scientific And Research Publications vol3,Issue2, Retrieved from <http://www.ijsrp.org/research-paper-0213/ijsrp-p14150.pdf>.

Table (1)

Describes metadata types.

Type	Description
Descriptive	It used to identify and describe the object. For example the title, subject, date, description, type of resource.
Structure	It is about how to Describe the descriptive metadata and determine what that attribute means. (chapter number, indexes, pages, sections, tables of contents)
Administrative	Provides information to help resource managing such as (when and who it was created, file type and other technical information.
Preservation	Is information that supports and documents the digital preservation process, and stores technical details on the format structure and use of the digital content (backup, refreshing cycles, migration..etc).
Technical	It is related to systems behaviour from business and functional specifications. (Accessibility, security issues, hardware and software documentation).

Table(2)

Shows Logs data metadata framework:

Structure	IP address (Internet Protocol User unique address)	URL (Uniform resource locator)	TimeStamp (Current time of an event recorded by computer)	Session ID (Unique identifier for a specific user)	Browsing Bytes (Transmitted bytes)	User agent string (Technical information)
Access Authority	<ul style="list-style-type: none"> • Url Path. • File source (Digital media, Printed stuff). • File types: Text, video, photo, Audio..etc • File location. • File size. • File description (Image resolution, dimensions ...etc). • When and who created. • Last modification date. 					
Rights Management	<ul style="list-style-type: none"> • Provenance. • Authenticity. • Preservation Activity. • Technical Environment. <ul style="list-style-type: none"> • Backup Plan, Preservation events. 					

	Technical	<ul style="list-style-type: none"> • Database Owner. • Physical characteristics (table space, extents, segments, blocks, partitions). • Performance tuning (processors, indexing). • Tables specifications and columns details. <small>Software specifications, versions and documentation.</small> • Constraints, the relationship between tables. • Data dictionary. • Compatibility. • Authentications and security issues e.g. formats, encryption keys, passwords. • Tracking of system response time. 					
	Descriptive	<ul style="list-style-type: none"> • String. • Not Null • Indexing 	<ul style="list-style-type: none"> • String. • Not Null • Indexing 	<ul style="list-style-type: none"> • Date Time. • Not Null • Indexing 	<ul style="list-style-type: none"> • Integer. • Not Null. • Indexing. • Unique. • Auto Increment 	<ul style="list-style-type: none"> • Float. • Not Null • Indexing 	<ul style="list-style-type: none"> • String. • Not Null • Indexing

Table(3)

Shows Metadata attributes and characteristics:

pue	Source of metadata.	<ul style="list-style-type: none"> • Internal metadata Generated by the creator. (File name, File format) • External metadata Created later by the someone else original creator (URLs, Rights and other legal information.
	Method of metadata creation.	<ul style="list-style-type: none"> • Automatic Metadata generated by a computer. (User transaction logs). • Manual metadata created by humans. (Descriptive metadata).
	Nature of metadata.	<ul style="list-style-type: none"> • Non-expert metadata Created by persons. • Expert metadata Created by experts.
	Status.	<ul style="list-style-type: none"> • Static. • Dynamic. • Long-term metadata. • Short-term metadata.
	Structure.	<ul style="list-style-type: none"> • Structured metadata (Predict standards). • Unstructured metadata (note fields, and other free text annotations).
	Semantics.	<ul style="list-style-type: none"> • Controlled metadata (standard vocabulary and authority). • Uncontrolled metadata.

Table(4)
illustrate webserver logs:

```
10.130.2.1,[30/Nov/2017:14:26:05,GET /css/normalize.css HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:26:05,GET /css/main.css HTTP/1.1,200
10.131.2.1,[30/Nov/2017:14:26:05,GET /css/style.css HTTP/1.1,200
10.131.0.1,[30/Nov/2017:14:26:05,GET /js/vendor/modernizr-2.8.3.min.js HTTP/1.1,200
10.130.2.1,[30/Nov/2017:14:26:06,GET /js/vendor/jquery-1.12.0.min.js HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:26:06,GET /bootstrap-3.3.7/js/bootstrap.min.js HTTP/1.1,200
10.130.2.1,[30/Nov/2017:14:26:07,GET /fonts/fontawesome-webfont.woff2?v=4.6.3 HTTP/1.1,200
10.130.2.1,[30/Nov/2017:14:26:12,GET /sign.php HTTP/1.1,200
10.131.0.1,[30/Nov/2017:14:26:49,POST /action.php HTTP/1.1,302
10.131.0.1,[30/Nov/2017:14:26:51,GET /login.php HTTP/1.1,200
10.131.0.1,[30/Nov/2017:14:27:00,POST /process.php HTTP/1.1,302
10.131.0.1,[30/Nov/2017:14:27:00,GET /home.php HTTP/1.1,200
10.131.0.1,[30/Nov/2017:14:27:01,GET /bootstrap-3.3.7/js/bootstrap.js HTTP/1.1,200

10.128.2.1,[30/Nov/2017:14:27:01,GET /js/vendor/moment.min.js HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:27:15,GET /contest.php HTTP/1.1,200
10.131.2.1,[30/Nov/2017:14:27:44,GET /contestproblem.php?name=RUET%20J%20Server%20Testing%20Contest HTTP/1.1,302
10.130.2.1,[30/Nov/2017:14:27:45,GET /countdown.php?name=RUET%20J%20Server%20Testing%20Contest HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:27:52,GET / HTTP/1.1,302
10.128.2.1,[30/Nov/2017:14:27:53,GET /login.php HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:27:53,GET /css/bootstrap.min.css HTTP/1.1,200
10.128.2.1,[30/Nov/2017:14:27:53,GET /css/font-awesome.min.css HTTP/1.1,200
10.131.2.1,[30/Nov/2017:14:27:53,GET /css/normalize.css HTTP/1.1,200
10.129.2.1,[30/Nov/2017:14:27:53,GET /css/style.css HTTP/1.1,200
```