



# AN APPROACH OF DYNAMIC FILTER FOR WEATHER ENVIRONMENT DEPENDENT ON TIME

<sup>1</sup>WA'EL JUM'AH AL-ZYADAT, <sup>2</sup>RODZIAH BINTI ATAN

<sup>1</sup>PhD Student, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor  
<sup>2</sup>Dr., Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, MALAYSIA.

E-mail: [wael\\_abade@yahoo.com](mailto:wael_abade@yahoo.com) , [rodziah@fsktm.upm.edu.my](mailto:rodziah@fsktm.upm.edu.my)

## ABSTRACT

In general dynamic filter has difficulty to execute the process because the Input /Output rapid change and its link-time directly almost work is synchronized within received data, the effectiveness for the results its and evaluations for all functions to process compatible usability, the structure designed for approach is Full Life Cycle Object-Oriented Testing (FLOOT) which aims the domain for weather its special environment which means high data sets is needed to support for end-user to predict the results in addition FLOOT give validity forward and backward processes and able to correct any error if discovered. Through experiments, we show our approach improves the performance of filtering information in two types of recognizes errors systems adopting the re-process data and visualization, respectively. In this paper will provide approach to resolve impact environment to data stream and enhancement filtering process interaction.

**Keywords:** *Data filtering, Enhanced Data, Clustering, Error Detecting, Full Life Cycle Object-Oriented Testing (FLOOT).*

## 1. INTRODUCTION

Information Filtering System had been referred to as time saving devices because the main goal of this information filtering is a prioritizing of information. Meaning, its seeking process in which documents are selected from a stream of incoming data to satisfy relatively stable and specific information needed, It also can take the form of highlighting items of high important or deleting items that are not considered relevant.

Current demand on data and database management are at high stake .latest technology of data management is a said to be based on service. To elaborate, those technologies are capable to store large amount of applications and data and provide services for public to access and use or work around stored applications for a fee. The faster they provide services, the more customers they will get around and will make more profit. [1] Recent emerging technology for example is Cloud Computing technology and environment. The emphasis is given to provide applied execution and

other computing services from a single point storage .another similar application example is Virtual PC technology which they provide services, application and execution data manipulation.

Environmental conditions at different locations on Earth can vary quite significantly from freezing cold and snowing at the poles to blisteringly hot and dry in the deserts. Between these two extremes there are both daily and seasonal variations. Winds blow from different directions, bright sunshine erupts from behind clouds and rain cascades from the sky. The weather data is reliable and is suitable to use intelligent database and data mining to trouble-free usefulness, The approach popup of Full Life Cycle Object-Oriented Testing (FLOOT) is used as it supports high quality product furthermore, easy feedback between stages can also be modified in any stages With the possibility of mistakes, almost non-existent This is one of the strongest and the best ways to use.

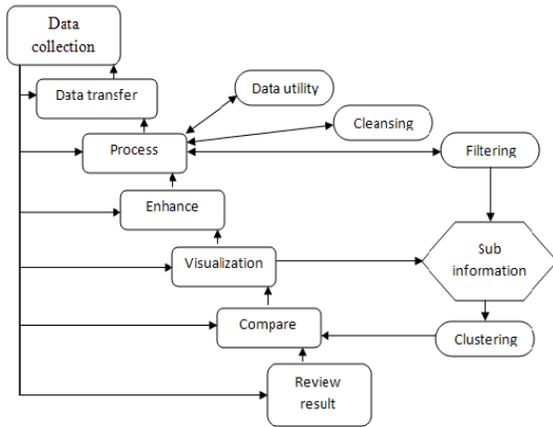


Figure 1: General approach dynamic filter weather data.

## 2. DATA COLLECTION (DC)

Its first stage is to accumulate data within the arena this means more specification for data. For missing data and complex data, structure will be made for them. The purpose of data collection is to obtain information to keep on record, to make decisions about important issues, and to pass information on to others. The data collection for weather in this cause is parallel with time to fetch the data. [2] The major problem during improvement of data collection times in many application such as area upload application is collect a large amount of data from several different hosts to a single destination in a wide area network.

This phase has two functions:

### 2.1. Scan

Make observation for the data and capture where it is located in space. This will persuaded the missing of the data and therefore the time to make declaration of location and what suitable area should be ignored [3].

Time =  $T$ , Data Scan =  $D_S$ , Queue =  $Q$ .  
 $F = T/D_S = 0$  then undeclared  $T=0$ ,  $D_S$  = capture data convert to binary.  
 $F = t/D_S =$  then binary data if  $T=1$  but static value then can simple to similarity in queue structure

TABLE I

STORE DATA AFTER COMPLETE DATA COLLECTION.

Q	D_S	T
00	0-1	1
01	5-14	1
02	15-24	1
03	25-34	1
04	35-44	1
05	45-54	1
06	55-64	1
07	65-74	1
08	75-84	1
09	85-94	1
10	95-104	1
11	105-114	1
12	115-124	1
13	125-134	1
14	135-144	1
15	145-154	1
16	155-164	1

The table contains three columns (Q, D\_S, T) it build depend in static function which mean the Time (T) =1 to easy simulative in structure and will less time complexity, Queue (Q) its main raw for major apply to structure data and in this time can recall for the data semi-structure, D\_s

### 2.2. Storage

This temporary function depends on the structure queue by logic. They contain missing data and trash. This will also take huge space in memory (use similar buffer technique). This function is divided: data is put in package to be repaired and in support of transfer, Asynchronies will be used.

## 3. DATA TRANSFER (DT)

This stage will carry the data such as buffer but the kind of semi-structure document use the XML to keep data from losing ; it is necessary because the type of data is parallel with time “real data in environment” regular of work to put data in frame for illustration SQL. Since majority of DATA ACCESS is filed in Databases Management, it transfers more functional facilities and freelance operation. [3] For this stage, many operation are included; similarity and sedulity such as using standard to remove the gap between Operating System and different software, [4] conclusion DT can depend on movement in bandwidth and finding correct schedule (move structure) with partition data in storage.[5] Due to data intensive application, many complicated scientific information are involved. Therefore, implementation of so called Data Transfer / Data



Grid is to simplify or to filter certain unused file to access.

#### 4. PROCESSES

This stage contains three main operations and its operation becomes strength relations where “synchronization” is used to explain mutual work.

##### 4.1. Data utility

[6] “Describing the value of a given data release is an analytical resource. This comprises the data’s analytical completeness and its analytical validity. Disclosure control methods usually have an adverse effect on data utility. Ideally, the goal of any disclosure control regime should be to maximize data utility whilst minimizing disclosure risk. In practice, disclosure control decisions is a trade-off between utility and disclosure risk”. [7] Data quality is linked with a strong data utility. Here lies the relative proximities to associate with the quality of the data selected; the measurement for data utility, its equal in many functional of data quality (e.g. Right data, in the Right Format structure or semi-structure, right time). The data utility has one functional usage to recognize between unknowable data and knowable data and to serve as the selection process between the two types prefer. Technique Information Retrieval (IR) will be employed.

##### 4.2. Data cleansing (DC)

This step is a continuous preprocesses. Usually, the first step in process is (missing data, normalization, etc...). [8] This process care for the raw data and how to make it quality data specifically to remove inside error data set, make corrections within format data and how to make validation check and how to put data in standards.[9] Cleansing RFID data needs an analyzing product lifecycle information in order to detect and correct errors across the sequence of fact. [10] The growth of data on the web makes information management knowledge discovery difficult, low quality and sometime contradictory data. By executing this data cleansing, it would be extremely helpful for user to identify the quality of the web user requested and in order for them to index more high-quality pages with limited resources.

The major functionality contains process:

-Error checking: the interest for this arena is to turn errors in all data set into percentage so if the error doesn’t make direct impact, the inside rule data can be moved to other functional. This means if the result is less than 10% it should be selected from the data set in buffer.

-Error Detecting: this function determine the location of error data or missing data depending on the present dataset for matrix.

-Error validation: its care designed for proof the error detecting its correct choose error and doesn’t skip any error data, and to make sure not to select meaningful data.

- Error Correction: due to repairing the errors e.g. duplicate records in this condition and time in case cant repair because of different time. It is limited work and fitness and it cannot be relied upon mainly.

##### 4.3. Data filtering

[11] This is concerning what other users or group want of database. In addition, there are two types of filters database; one is Horizontal filters. The main concept labor depends on row for mirror and source. Second, Vertical filters filter columns of data mirror and source, everyone imagines the kind of database and what is the parameters and structure for database. [12] Usually the filter operation will ensure the database clean and clear for users or group otherwise it will need Information Retrieval (IR) to fix database. But this approach will give users the chance to repair itself because the feedback is supported with a fix which is possible via FLOOT. The information filter has many techniques for this Scope. Collaborative Filtering (CF) is preferred because [13] it is suitable for large database and accept high modified data (e.g. Update item, add item, etc..). In addition, it ranks from reliable sources, use wide population, advisors and weight sum; [14] The first system supporting CF is its TAPESTRY for Email message which functions is to construct filtering. It has two steps; first level improve the filtering function, classify and use Boolean function to select pure information. It could be accepted or reject below conditions without exception. In this case, a link among Data cleansing (DC) and series of stages show every high relation amid stages. This level will explain how the sequenced reaction data process within

come together. But it also depends on structure types (structure, semi-structure) upon the direct way. After reaching the target by default, in the second level, which also means the filter arrives at what the user wants, the fault user or group to reach target is likely to overlook in certain areas. Usually, the result rejects because the wrong transfer of database to A. Although the process is correct but sometimes we are unable to retrieve the database due to the user choosing the wrong data.

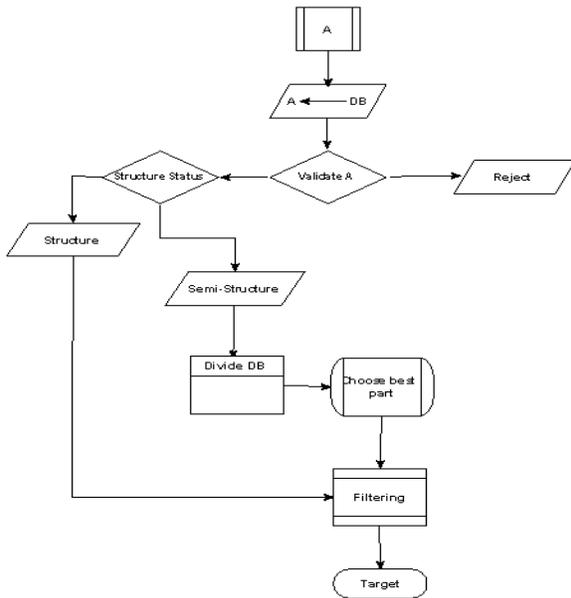


Figure 2: formation filtering data until to target.

The above figure shows the initialization outfit the variable A to receive all databases contents within details after that validates it to confirm that it did not loss any data. If any losing data is found, will stop execute and return to process. But success validated data will move to the next function to determinate the status, the structure and to choose duration step. Table 2 traces formation filtering and how it reaches to target and which time stop it stops.

TABLE 2: EXPLAIN FORMATION STEPS.

	A	A ← DB	Validate A	Structure	Semi-Structure	Divide DB	Choose best part	Filtering	Target	Reject
1	●	●	●	●				●	●	
2	●	●	●		●	●	●	●	●	
3	●	●	●							●

The above table has three Cases illustrating the results, the critical phase Validate A to continue or to attain target or reject or drop unstructured database because previously there was one stage responsible to remove or repair the unstructured databases.

5. ENHANCED DATA

It is known as a backward-compatible which allows improved data transmission rates. It is an extension on top and it is an upgrade that provides more increase in both the capacity and performance of real data, it does this by introducing sophisticated method of coding and transmitting data and higher delivering higher. Current database strongly depends on the amount of available memory, be it as I/O buffers or as work space for query processing algorithms, it more clear to explain where is it suitable applied. The SQL closely simple JOIN query as commonly used in relational DB system [15]. The pour enhance data below Data Quality because it has at least four combined features (Correct data, Standardize data, Verify data, Reliability and Trust).

Visualization [[16] [17]] is any technique for creating images, diagrams, or animations to communicate a message. Visualization through visual imagery has been an effective way to communicate both abstract and concrete ideas. For examples the weather data; their high impact with environment and direct link with environment its involve at least use two-dimensions, or three-dimensions but prefers to use three dimensions to make high resolution and more specifically as well as more facilities and usability for comparison to do superior related.

6. SUB-INFORMATION

It is a kind of data validation. Taking the captured data after filtering process also capturing the same data in visualization phase to make improvement if have any fault during process or wrong data to construct repair before final review to user or group. In the case of fault, exception handling will be used to make backward function. This approach supports one framework. Unfortunately if backward is used, then it will be wasting time and needs a long time and complex.



## 7. CLUSTERING

[18] [19] (It's crucial for many applications to evaluate the results within usage clustering algorithm and do measurement of data sets dynamic or static; however it can be absorbed by historical data which has high modify and compatible to patterns and also large amount data arriving in real time. It easily repairs the requirement for comparing.

## 8. COMPARE

This phase is making the process effective and determine quality and quantity of data based on the technique Clustering "artificial intelligent". These phases can measure the test's results and fail percentage.

## 9. REVIEW RESULT

The outcome for the results and determine the profile data in this phase, can recall for the data information because it is built-in.

## 10. CONCLUSION

In this paper, we have investigated the problem of dynamic Information Filtering with the aim of identifying the relevant high relational between re-process and visualization to save content data within regular structure and improve a resolution information, clustering its direct effect to filtering utilization. We ended future work to recommended from visualization synchronization with filtering information to make core correctness comparison to shall quality and validity information as well as clear appearance errors to stakeholder or item .

## 11. ACKNOWLEDGEMENTS

This paper determination dataset scope Malaysia weather climate from 2008 and 2009 and will cover 13 States (Alor Setar, Bayan Lepas, Cameron Highlands, Chuping, Datalist, Ipoh, Kota Bharu, Kuala Terengganu Airport, Kuantan, Malacca, Mersing, Petaling Jaya, Senai).The data contract from MALAYSIA METEOROLOGICAL DEPARTMENT (MMD).

## REFERENCES:

- [1] Pearson, S. *Taking Account of Privacy when Designing Cloud Computing Services*. IEEE. 2009:9.
- [2] Yiqun Liu, C.W., Min Zhang, Shaoping Ma. *Web Data Cleansing for Information Retrieval using Key*.
- [3] Leeser, S.T.a.M. *The DT-Model: High-Level Synthesis Using Data Transfers*. IEEE. 1998:4.
- [4] Mahesh Visvanathan\* Adagarla, B.S.G., H Lushington Peter Smith. *Cluster Validation: An Integrative Method for Cluster Analysis*. IEEE. 2009:5.
- [5] Bill Allcock1 Joe Bester1 John Bresnahan1 Ann L. Chervenak2 Ian Foster1; Tueckel1, C.K.S.M.V.N.D.Q.S. *Data Management and Transfer in High-Performance Computational Grid Environments*. University of Southern California.24.
- [6]IBM. *what is the data filter*; [modified 2005]. Available from: <http://publib.boulder.ibm.com/infocenter/weahelp/5.1/index.jsp?topic=/com.ibm.websphere.db2e.doc/dbessc0910.html>
- [7] Zhang, M.J.S.a.L. *Comparison of Cluster Representations from Partial Second- to Full Fourth-Order Cross Moments for Data Stream Clustering*. IEEE. 2008:10.
- [8] Colby, J.R.S.D.H.T.L.S. *A Deferred Cleansing Method for RFID Data Analytics*. ACM. 2006:12.
- [9] guzzi, A. *Visualizing scientific contributions*. University della svizzera italiana.34. Proceedings of the 32nd international conference on Very large data bases,( Pages: 175 - 186 ,year: 2006:12)
- [10] Yiqun Liu, C.W., Min Zhang, Shaoping Ma. *Web Data Cleansing for Information Retrieval using Key*. Special interest tracks and posters of the 14th international conference on World Wide Web(Pages: 1136 - 1137 ,Year of Publication: 2005,ISBN:1-59593-051-5).
- [11] Laschov, D. *Collaborative Filtering*; [modified 2006; cited 2009]. Available from: [http://74.125.153.132/search?q=cache:av\\_I5I\\_a1MgJ:bakara.eng.tau.ac.il/~semcomm/slides7/dima.pps+dima+laschov&cd=1&hl=en&ct=clnk&gl=my&client=firefox-a](http://74.125.153.132/search?q=cache:av_I5I_a1MgJ:bakara.eng.tau.ac.il/~semcomm/slides7/dima.pps+dima+laschov&cd=1&hl=en&ct=clnk&gl=my&client=firefox-a)
- [12] Rappoport, R.R.A. *The NVI Clustering Evaluation Measure*. ACM. International Conference On Computational Linguistics



- (Year of Publication: 2009,ISBN:978-1-932432-29-9).
- [13] [modified 2009; cited 2009 26/12/2009]. Available from: <http://web4.cs.ucl.ac.uk/staff/jun.wang/blog/topics/research-resources/collaborative-filtering/>
- [14] Wang's, D.J. *Collaborative Filtering Explained*
- [15] William C. Cheng, C.-F.C., Leana Golubchik, Member, IEEE, Samir Khuller, and Yung-Chun (Justin) Wan. *A Coordinated Data Collection Approach: Design, Evaluation, and Comparison*. IEEE. 2004;22:15
- [16] Zhou, X. *Data Quality: A Database Perspective*.
- [17] Scofield, M. Michael Scofield. *The Data Administration Newsletter*, LLC (TDAN.com) 2005.
- [18] Kossmann, P.M.F.D. *Quality of Service in Stateful Information Filters*. ACM. 2006:6.
- [19] Shapiro, G.G.L.D. *Data Compression and Database Performance*. ACM/IEEE-CS Symp. 1991:10.
- [20] Join Queries. Available from: [http://www.oracle-dba-online.com/sql/join\\_queries.htm](http://www.oracle-dba-online.com/sql/join_queries.htm)
- [21] Alena, C.L.R.E.H.D.M.H.R.L. *Using Fuzzy Clustering for Real-time Space Flight Safety*. Nasa Research.5.
- [22] Baudisch, P. *Dynamic Information Filtering [Dissertation]*. [Darmstadt]: Darmstadt; 2001. 3 p.
- [23] Chapman, A. *PRINCIPLES AND METHODS OF DATA CLEANING*. Australian Biodiversity Information Services. 2005:75.
- [24] Daniel Crabtree, X.G., Peter Andraea. *Improving Web Clustering by Cluster Selection*. IEEE. 2005:7.
- [25] Kempfert, W.S.a.K. *Applications of Rapid Scan Data Collection*
- [26] Nicolet Spectroscopy Research Center.AN-9584:2.
- [27] Lee, M.J. *Toxics Release Inventory - An Environmental Database*. IASSIST. 2000:2.
- [28] NAKAMURA, K.Y.a.Y. *Dynamics Filter -- Concept and Implementation of On-Line Motion Generator for Human Figures*. IEEE; 2000. p. 8.
- [29] Netherlands, S. DATA UTILITY; [modified 2005]. Available from: <http://stats.oecd.org/glossary/detail.asp?ID=6905>
- [30] Ng, A.A.Y.-K. *A dynamic attribute-based data filtering and recovery scheme for web information processing* ACM. 2008 2007:9.
- [31] Shankaranarayanan, A.E.a.G. *Utility-Driven Assessment of Data Quality*. ACM. 2007 May 2007;38:19.