The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Mining RDF (Linked Data) using Éclat algorithm

Wria Mohammed Salih Mohammed[1], Payman Othman Rahim[2]

Department of Computer Science, College of Science, University of Slemani, Slemani, Iraq[1,2]

payman.rahem@univsul.edu.iq[2]

Abstract:

Basket market analysis is one of the most widely used groups of data mining and have been extensively utilized for analyzing data to extract interesting information from huge amount of data. Also, Studies over the past two decades have provided important information on semantic web as it is part of (World Wide Web Consortium) W3C. Both data mining and semantic web have several key features to mine semi-structured dataset and having an accurate result. The methodological approach taken in this study is combining both Éclat algorithm and RDF (Resource Description Framework) dataset based on the process of converting RDF into dataset and mining it. Firstly, RDF data is checked for validation, and then it needs to convert into traditional dataset. This process requests SPARQL as a query language. Thus, it needs to imply Éclat (Equivalence class Transformation) algorithm on traditional dataset. This experiment illustrates that semantic web and data mining have significant results in mining semi-structured dataset. This paper hands out how mixing RDF and Éclat algorithm is influent. For this technique different data source can be used, however, for this paper particularly products in a supermarket are going to use as a main dataset.

Keywords:RDF, Data Mining, Linked Data, Éclat, SPARQL.

الملخص:

يعتبر تحليل سلة التسوق واحدا من اكثر التقنيات استخداما في تعريف البيانات وقد كان و مايزال الاكثر استعمالا في تحليل البيانات لاستخلاص معلومات هامة من كم هائل من البيانات. ولقد زودتنا الدراسات خلال السنوات الماضية بمعلومات هامة عن شبكة السيمانتك كونها جزءا من جمعية الشبكات العالمية. و تمتلك كلا من تعريف البيانات و شبكة السيمانتك خصائص عديدة لتعريف ملف البيانات الانشائية الجزئية و الحصول على نتائج دقيقة و صحيحة. النهج المتبع في هذه الدراسة عبارة عن دمج نظام (ئيكلات) للحلول الحسابية ونظام اطار وصف الموارد (آر دي اف) لتحليل ملف البيانات بناءا على عملية تحويل (الآر دي اف) الى ملف بيانات و تعريفها. اولا, يتم فحص بيانات (الآر دي اف) لغرض تأكيد فاعليتها, ثم يجب تحويلها الى ملف البيانات التقليدية. هذه العملية تتطلب لغة سباركل كلغة الاستعلام. وهكذا تحتاج لتطبيق نظام (ئيكلات) للحلول الحسابية على الملفات التقليدية. توضح هذه التجربة بان لكل من شبكة السيمانتك و تعريف البيانات نتائج مهمة في تعريف ملف البيانات الانشائية الجزئية. يبين هذا البحث بان دمج (الآر دي اف) و(ئيكلات) للحلول الحسابية يتم بشكل مؤثر. ويمكن استخدام موارد معلوماتية لهذه التقنية, وفي هذه الدراسة تم استخدام منتجات احدى الاسواق (السوبرماركت) كملف بيانات.

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

پوختە:

شیکردنەوەی سەبەتەی بازارِکردن یەکیکە لە تەکنیکە هەرە باوەکانی بەکارهینان لە ناساندنی داتاکان وئێستاش زۆرترین بەکارهێنانی تۆمار کردووە لە شیکردنەوەی داتاکان بۆ بەدەستهێنانی زانیاری گرنگ لە بڕێکی زۆر داتا. توێژینەوکانی ساڵانی پێشوو زانیاری زۆر گرنگمان دەخەنە بەر دەست سەبارەت بە (تۆڕی سیمانتیک) کە بەشێکە لە کۆمەڵەی تۆڕە جیهانیەکان. ناساندنی داتاکان و (تۆڕی سیمانتیک) چەند خەسڵەتێکیان هەیە بۆ ناساندنی فایلی داتا نیمچە دروستکارییەکان و دەستکەوتنی ئەنجامی ڕاست ودروست. ڕێبازی پەیڕەوکراو لەم توێژینەوەیەدا بریتیە لە تیکەڵ کردنی سیستمی (ئیکلات) بۆ چارەسەرە ماتماتیکیەکان و سیستمی چوارچێوەی وەسفی سەرچاوەکان کە ناسراوە بە (ئار دی ئێف) بۆ شیکردنەوەی فایلی داتاکان لەسەر بنەمای گۆڕینی (ئار دی ئێف) بۆ فایلی داتاو ناساندنی. سەرەتا شیکاری بۆ داتاکانی (ئار دی ئێف) ئەنجام دەدرێت بۆ دڵنیابون لە کاریگەری داتاکان و دواتر گۆڕینی بۆفایلی داتای تەقلیدی. ئەم کرداره پێویستی بە زمانی (سپارکڵە). بەم شێوەیە پێویستی بە جێبەجێ کردنی سیستمی (ئیکلات) بۆ چارەسەرە ماتماتیکیەکان بۆسەرفایلە تەقلیدییەکان. ئەم ئەزموونە ئەوەمان بۆ دەردەخات کەوا هەر یەک لە (تۆڕی سیمانتیک) و ناساندنی داتاکان ئەنجامی گرنگیان هەیە لە ناساندنی فایلی داتا نیمچە دروستکارییەکان. ئەم توێژینەوەیە دەریدەخات کەوا تێکەڵکردنی سیستمی (ئیکلات) بۆ چارەسەرە ماتماتیکیەکان و سیستمی چوارچێوەی وەسفی سەرچاوەکان کە ناسراوە بە (ئار دی ئێف) بە شێوەیەکی کاریگەر ئەنجام دەدرێت. لەم توێژینەوەیەدا بەرهەمی یەکێک لە سوپەرمارکێتەکان بەکارهاتووە وەک فایلی داتاکان.

1- Introduction

Information on the web has been astonishing increased, RDF(Resource Description Framework) data is one of the well-known ways to interchange data on the web, also, it is part of w3C's Semantic web activity (DECKER, et al., 2000). Furthermore, it is originally written based on xml format, and it is written in a way that Computer can read and understand it, however it is not designed to display to people. RDF is mostly presented in the structure of subject, predicate, and object (AbedjanEmail & Naumann, 2013). RDF has become an important to be mined using different techniques, as it is a semi-structured data. it has much attention by semantic web and database communities because it is used to interchange and linking data over the world wide web.

Moreover, Data mining also named Knowledge Discovery from Data (KDD), data mining is a sequence technique to find useful and previously unknown knowledge or patterns from huge data. There are diverse data sources that can be mined, such as databases, the web, data warehouses, or any other data sources that capable to be mined (Han, Pei, & Kamber, 2012). In this paper, a data mining technique will be used on RDF data. there are many algorithms in data mining including classification, clustering or Association Rule Mining algorithms, ARM is used in this paper, to be more particularly Éclat algorithm is going to be used, because, Éclat can find and analysis important information from transactions and comparing various items in the same transaction. Apriori and Éclat are most frequent item set mining algorithms. Éclat is an unsupervised learning algorithm that is a part of association rule mining. Furthermore, SPARQL is definitely not designed to query relational data, however, it utilizes to query RDF data, working with semantic web without SPARQL is like to operate with relational database without SQL (DuCharme, 2013).

DOI: http://dx.doi.org/10.25098/3.2.21

2- Related work

Over the past two decades, many researches had published in area of data mining and semantic web, the basic of semantic web data is RDF which is based on XML format. it can be seen that there are many researches about data mining and RDF data. In (AbedjanEmail & Naumann, 2013) , the authors shows rule-based techniques to suggest predicate, enrich data, improve ontology, relax queries, however, they have problem with missing data when it is enriched and inconsistencies in suggesting predicated. The paper (AbedjanEmail & Naumann, 2013), illustrates the benefits to integrate association rule mining with RDF data. This paper (Borgelt, 2012), presents an introduction of frequent item set mining, they continue how structure the search space to avoid redundant search. They pruned a priori property and showing the reducing output. This paper (Borgelt, 2012), categorized to the following steps:

- Association rules mining to develop algorithm.
- Apply data mining tools to application area.
- Using association rule mining to technologies.

Another research that published was about comparison between of re-description mining and association rule mining for finding definition of class in linked open data (ReynaudEmail, Toussaint, & Napoli, 2019). In this paper, the basic of re-description mining and make precise the principles of definition discovery. They extract datasets from DBpedia, as a result different output analyzed related to re-description mining and association rule mining application. There is another paper (Baratia, Baia, & Liub, 2017) is about ignoring knowledge encoded at the schema-level negatively impacts the interpretation of discovered rules. They introduced an approach about mining semantic web using rule mining, they compare rules found from SWRAM with some techniques. Furthermore, another paper that illustrates data mining system for linked data is (Venkata, Kappara, & Ichise, 2011), they proposed a model to help inter-act with linked data to make structured web data. they used various data source for analysis and perform data mining to obtain the result.

3- Preliminaries

This section briefly introduces the techniques that used in this research:

3-1. RDF:

RDF is basic of XML, which can be exchanged among application or users. RDF was invented by W3C in 1999 (Mohammed & Saraee, 2016).

RDF is a part of semantic web data which is graph-based data. it is attempt to solve problem of URIs that integrated. Using RDF combination with data is faster and more robust than traditional in modification of schema. RDF is a main key for linked data. linked data is well-known published data on the world wide web. It is invented by Tim Berners-Lee (Gayo, Prud'hommeaux, Boneva, & Kontokostas, 2018). It is basically utilized to move machine-understandable data on the web. Figure 1 shows the structure of RDF data which has three elements (subject, predicate, and object).

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.w3.org/computer_Scinece/stage_1">
    <dc:course>Structure Programming</dc:course>
    <dc:course>Logic Design</dc:course>
    <dc:course>Computer Organization</dc:course>
    <dc:course>Database</dc:course>
   </rdf:Description>
</rdf:RDF>
```

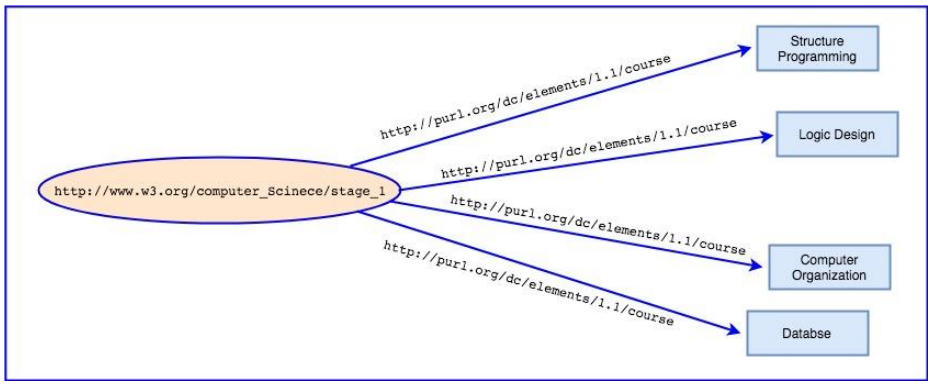


Figure 1: RDF structure

3-2. SPARQL

When the user wants to query RDF, they need to use SPARQL, SPARQL is W3C recommendation and it is one of the well-known query languages in semantic web. SQL is query language for relational database, but SPARQL is for semantic web (Mohammed & Saraee, 2016). Furthermore, RDF is not data structured, RDF consists of three parts: subject, predicate and object, and SPARQL is working according to these three parts (DuCharme, 2013). Figure 2 is an example about SPARQL which written on the RDF from figure 1.

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

```
select ?course
where
{
<http://www.w3.org/computer_Scinece/stage_1>
<http://purl.org/dc/elements/1.1/course>
?course.
}
order by ASC(? course)
```

Course
Structure Programming
Logic Design
Computer Organization
Database

Figure 2: SPARQL

3-3. Association Rule Mining

One of the most popular data mining techniques is Association Rule Mining (ARM), it also called Market Basket Analysis, ARM is introduced by Agrawal, Imielinski and Swami in 1993, the main goal of using ARM is to extract important association, frequent patterns, correlations or casual structures among set item in transaction database or any other sources (Koh & Rountree, 2010).

If X⇒ Y, in a database, tuples satisfying that X are likely to satisfy Y. a well-known ARM algorithm is Priory algorithm and Éclat as well. ARM uses two rule factors: support and confidence. Also, ARM can be used in medical diagnosis, Marketing, decision making and business management (Shimada, Hirasawa, & Hu, 2006). The result of ARM is set of rules such as (98% of all customer in super market, they buy product A and B also buy product C) as shows in Figure 3. These rules are beneficial for cross-marketing, catalog design, product placement and customer segmentation. (Hidber, 1999) . Frequent patterns are widely used in ARM; it is the frequently of having two items in a transaction. For example, a set of product, such as banana and apple that present frequently together in one transaction data set is frequent item set. A sequence such as buying first a mobile, then a charger, and then a Network cable, if it happens frequently in a history of supermarket database, is (frequent) sequential pattern (Han, Pei, & Kamber, 2012). In this research, Éclat algorithm is used, which is explained in the next section.

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

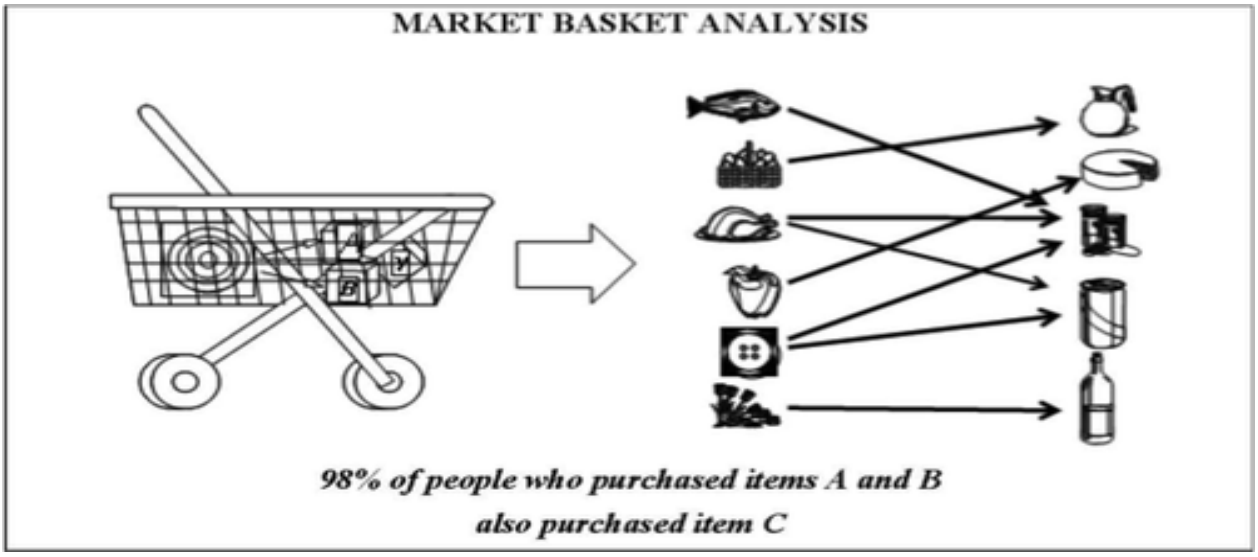


Figure 3: ARM

3-4. Éclat

Originally, the Éclat algorithm is used inside the Apriori algorithm, when the database is stored in the vertical layout, the support of a set can be counted much easier by simply intersecting the covers of two of its subsets that together give the set itself (Maimon & Rokach, 2005). Moreover, Éclat stands for Equivalence class Transformation. It transforms dataset from horizontal data format into vertical data format. It is simply mined the transformed data (Han, Pei, & Kamber, 2012). The database scans once only using Éclat algorithm, support is counted in this algorithm. However, Confidence is not counted using Éclat (kaur & Grag, 2014). The steps of the algorithm are explained in the followings and the execution of the algorithm is shown in Figure 4. The most useful point of using Éclat algorithm is to obtain the rate of the support for each item according to dataset. Another reason of using éclat is easy to implement and fast to having results.

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Algorithm *Éclat* (*FP*, support: s)

Begin

For each *Pi* ∈ *FP* do

Begin

FPi = {}

For each *Pj* ∈ *FP*, such that j> i do

Begin

Pij = *Pi* U *Pj*

Tideset (*Pij*) = Tideset (*Pi*) ∪tidset (*Pj*)

Support(*Pij*) = |tidset(*Pij*)|

If(support(*Pij*) ≥ s)

Add *Pij* to *FPij*

End

Eclat (*FPi*,s)

End

End (Aggarwal & Han, 2014)

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

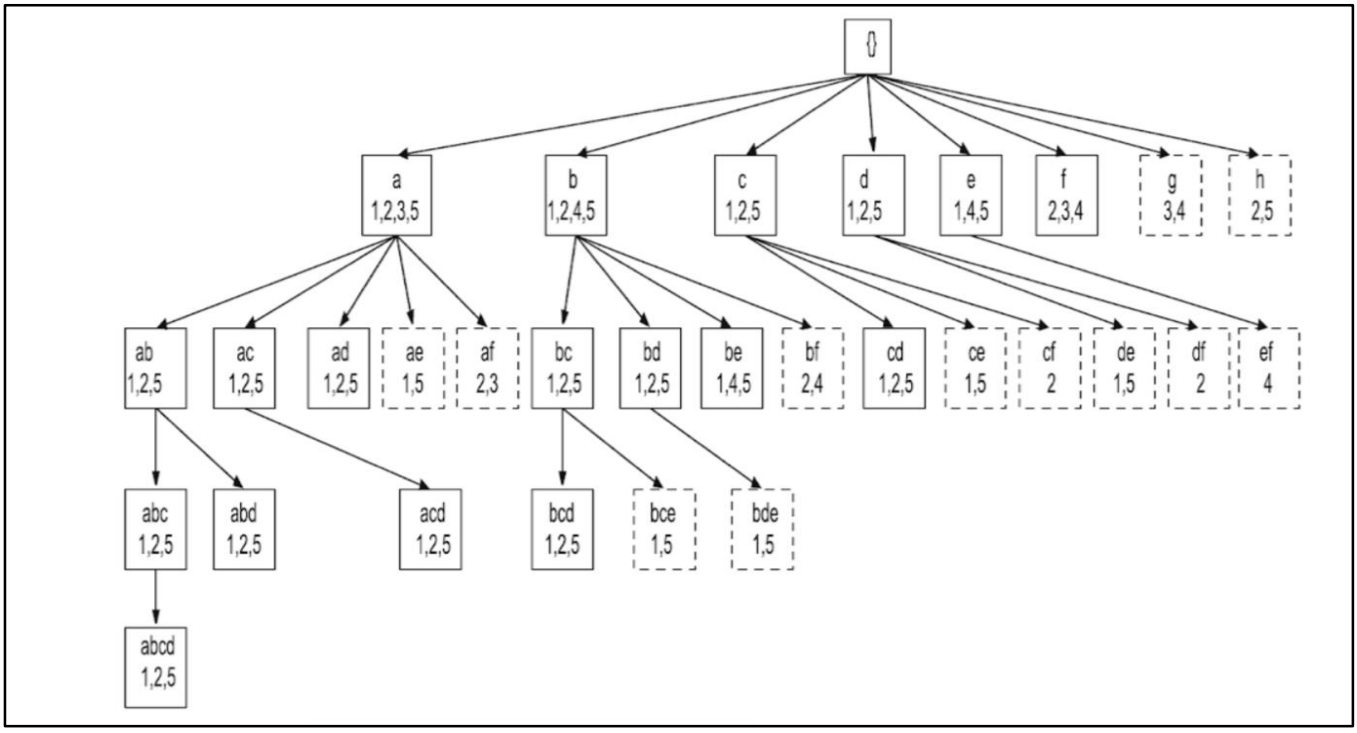


Figure 4: Execution of Eclat (Aggarwal & Han, 2014)

There is an example which is about Éclat algorithm

Table 1: item set

Transaction id	**Items**
T100	{10, 30, 40}
T200	{20, 30, 50}
T300	{10, 20, 30, 50}
T400	{20, 50}
T500	{10, 20, 30, 50}

The **support** of the item set is the number of appearing item set in the transaction database. The item set has to change to vertically.

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Table 2: support of items

Item sets	**support**
{10}	3
{20}	4
{30}	4
{50}	4
{10, 20}	2
{10, 30}	3
{10, 50}	2
{20, 30}	3
{20, 50}	4
{30, 50}	3
{10, 20, 30}	2
{10, 20, 50}	2
{10, 30, 50}	2
{20, 30, 50}	3
{10, 20, 30, 50}	2

The {20, 30, 50} item set shows in transactions T200, T300 and T500, it means the item set has three supports.

4. Problem statement

The problem of this paper is achieving on of the most famous association rule mining algorithm on RDF data to detect important information from huge amount of data from various data sources, because there is not such an implementation to apply Éclat algorithm on RDF data. Éclat is unsupervised approach to change data from horizontal to vertical, then mine the dataset vertically. This algorithm finds the relationship among data items in a transaction or dataset. There are some examples of Éclat algorithms:

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

- Using Éclat to Basket market analysis
- Using Éclat to profile analysis.
-

Moreover, in this project SPARQL has been used as a query language on RDF data. another problem is that it is complex to analysis RDF data without SPARQL because there is not an adequate way to convert unstructured or semi-structured dataset into traditional structured dataset without SPARQL.

The main goal of combining both area: RDF data and Éclat algorithm is to improve the result of mining by gathering data from unstructured or semi-structured data sources. Also, the purpose of using Éclat is to change data from horizontal to vertical then apply frequent item generation. This algorithm is important to find the relations among data items in the same dataset.

5. Methodology:

The methodology of this research consists of three different steps to obtain the results which shows in Figure 5. This project mixes both RDF data and Éclat algorithm. In the first step, semi-structured dataset (RDF dataset) are going to use, the complicity of the RDF does not allow to apply mining directly, but it needs to convert the RDF data into Traditional dataset. Next, RDF dataset is going to convert into Traditional dataset using SPARQL techniques. Finally, the converted dataset is required to mine using Éclat algorithm.

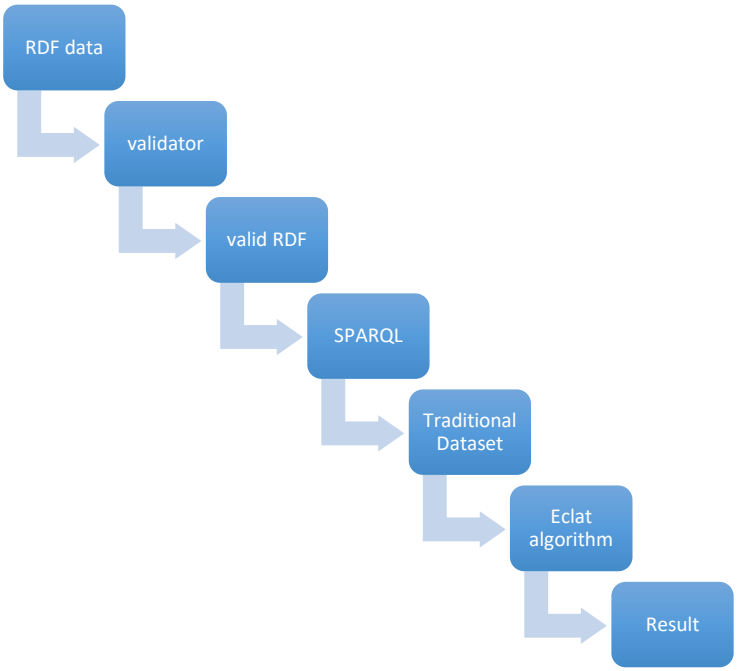


Figure 5: Steps RDF Mining using Elcat

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

these steps are going to explain in the followings:

6. results

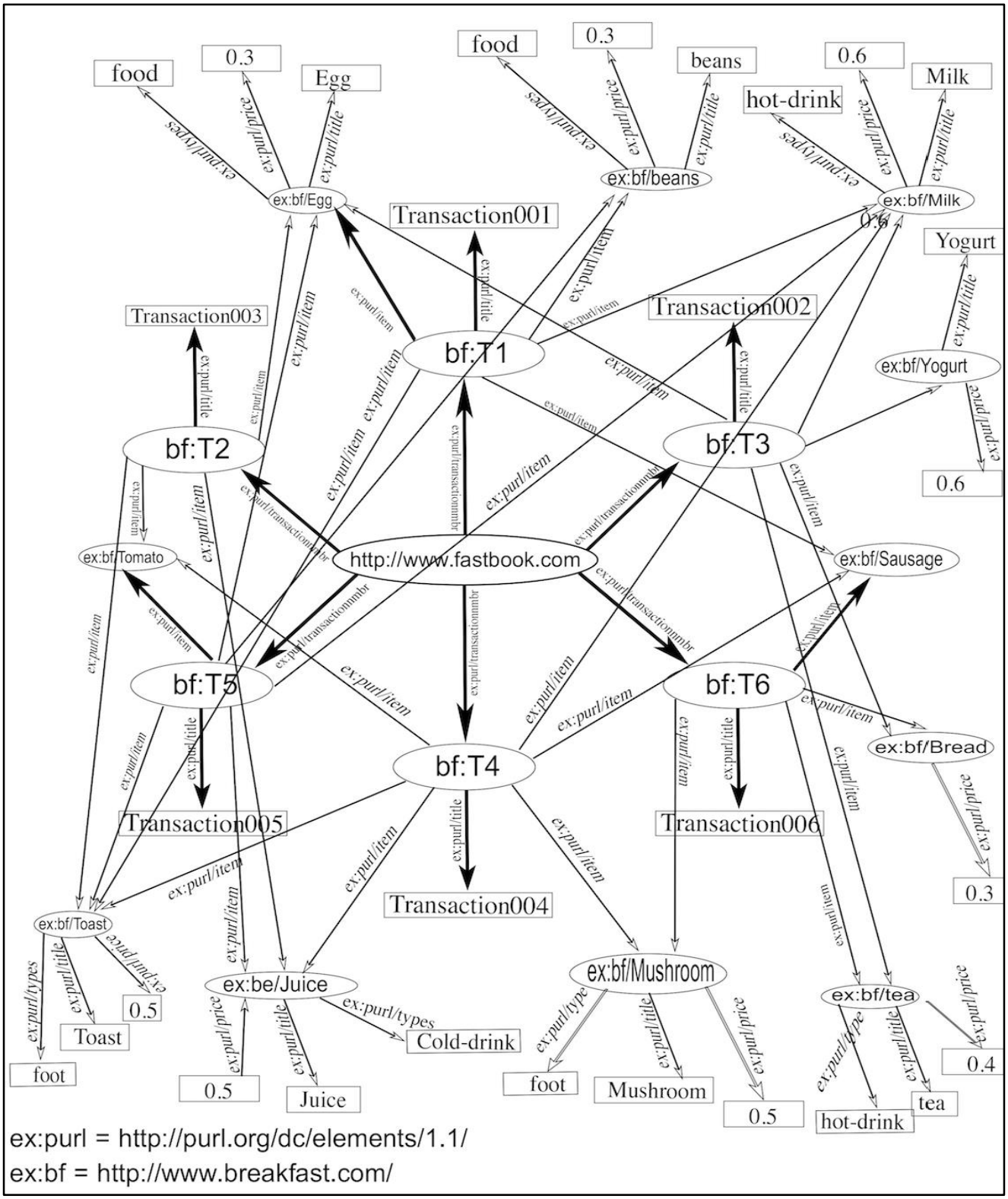
6-1. RDF Dataset

In this paper, semi-structured (RDF) dataset are going to use, RDF consists of three different parts (subject, predicate, object). These three components are called Triples (as shown in Figure 7) and it is beneficial for mining. The RDF can be shown in graphic which appearing all subjects, predicates and objects as shown in Figure 6. There is a sample of RDF data.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.breakfast.com/">
    <dc:transactionnumbr>
      <rdf:Description rdf:about="http://www.breakfast.com/T1">
        <dc:title>Transaction001</dc:title>
      </rdf:Description>
    </dc:transactionnumbr>
    <dc:transactionnumbr>
      <rdf:Description rdf:about="http://www.breakfast.com/T2">
        <dc:title rdf:parseType="Literal">Transaction002</dc:title>
      </rdf:Description>
    </dc:transactionnumbr>
    <dc:transactionnumbr>
      <rdf:Description rdf:about="http://www.breakfast.com/T3">
        <dc:title>Transaction003</dc:title>
      </rdf:Description>
    </dc:transactionnumbr>
    <dc:transactionnumbr>
      <rdf:Description rdf:about="http://www.breakfast.com/T4">
        <dc:title>Transaction004</dc:title>
      </rdf:Description>
    </dc:transactionnumbr>
```

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

In the above example, there is an RDF data which includes number of transactions with the items, each item has price, title and type. RDF is used in this paper, because RDF is unstructured dataset. It



is not like a normal dataset to work on it.

Figure 6: RDF Graph

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

In figure 6, the structured of RDF was explained. RDF consists of three different parts (subject, predicate and object), for instance (http://www.breakfast.com/) is subject and (http://purl.org/dc/elements/1.1/transactionnumbr) is predicate and (http://www.breakfast.com/T1/) is object. As mentioned before object can be URI or literal but predicate and subject cannot be literal, for example, (Transaction001) is not a URI, it is a literal only. Also, this is the same for figure 7.

Subject	Predicate	Object
http://www.breakfast.com/	http://purl.org/dc/elements/1.1/transactionnumbr	http://www.breakfast.com/T1
http://www.breakfast.com/T1	http://purl.org/dc/elements/1.1/title	"Transaction001"
http://www.breakfast.com/	http://purl.org/dc/elements/1.1/transactionnumbr	http://www.breakfast.com/T2
http://www.breakfast.com/T2	http://purl.org/dc/elements/1.1/title	"Transaction002"^^http://www.
http://www.breakfast.com/	http://purl.org/dc/elements/1.1/transactionnumbr	http://www.breakfast.com/T3
http://www.breakfast.com/T3	http://purl.org/dc/elements/1.1/title	"Transaction003"
http://www.breakfast.com/	http://purl.org/dc/elements/1.1/transactionnumbr	http://www.breakfast.com/T4
http://www.breakfast.com/T4	http://purl.org/dc/elements/1.1/title	"Transaction004"
http://www.breakfast.com/	http://purl.org/dc/elements/1.1/transactionnumbr	http://www.breakfast.com/T5
http://www.breakfast.com/T5	http://purl.org/dc/elements/1.1/title	"Transaction005"
http://www.breakfast.com/	http://purl.org/dc/elements/1.1/transactionnumbr	http://www.breakfast.com/T6
http://www.breakfast.com/T6	http://purl.org/dc/elements/1.1/title	"Transaction006"
http://www.breakfast.com/toast	http://purl.org/dc/elements/1.1/title	"toast"
http://www.breakfast.com/toast	http://purl.org/dc/elements/1.1/price	"0.4"
http://www.breakfast.com/toast	http://purl.org/dc/elements/1.1/types	"food"
http://www.breakfast.com/egg	http://purl.org/dc/elements/1.1/title	"egg"
http://www.breakfast.com/egg	http://purl.org/dc/elements/1.1/price	"0.3"
http://www.breakfast.com/egg	http://purl.org/dc/elements/1.1/types	"food"
http://www.breakfast.com/sausage	http://purl.org/dc/elements/1.1/title	"sausage"
http://www.breakfast.com/sausage	http://purl.org/dc/elements/1.1/price	"0.7"
http://www.breakfast.com/sausage	http://purl.org/dc/elements/1.1/types	"food"
http://www.breakfast.com/beans	http://purl.org/dc/elements/1.1/title	"beans"
http://www.breakfast.com/beans	http://purl.org/dc/elements/1.1/price	"0.5"
http://www.breakfast.com/beans	http://purl.org/dc/elements/1.1/types	"food"
http://www.breakfast.com/tomato	http://purl.org/dc/elements/1.1/title	"tomato"
http://www.breakfast.com/tomato	http://purl.org/dc/elements/1.1/price	"0.1"
http://www.breakfast.com/tomato	http://purl.org/dc/elements/1.1/types	"food"
http://www.breakfast.com/juice	http://purl.org/dc/elements/1.1/title	"juice"
http://www.breakfast.com/juice	http://purl.org/dc/elements/1.1/price	"0.5"
http://www.breakfast.com/juice	http://purl.org/dc/elements/1.1/types	"cold-drink"
http://www.breakfast.com/bread	http://purl.org/dc/elements/1.1/title	"bread"

Figure 7: Triples of Product dataset

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

6-2. Convert RDF Dataset into Traditional Dataset

In this part, RDF dataset is going to convert into traditional dataset. SPARQL is a query language for querying on RDF data, there is a sample of SPARQL to retrieve dataset as traditional dataset.

```
select ?transactionTitle,?title,?price,?types
where
{
<http://www.breakfast.com/><http://purl.org/dc/elements/1.1/transactionnumbr> ?transaction.
?transaction <http://purl.org/dc/elements/1.1/title> ?transactionTitle.
?transaction <http://purl.org/dc/elements/1.1/item> ?item.
?item <http://purl.org/dc/elements/1.1/title> ?title.
?item <http://purl.org/dc/elements/1.1/types> ?types.
?item <http://purl.org/dc/elements/1.1/price> ?price.
}
order by ASC(?transaction)
```

According to the SPARQL, which is written just above, SPARQL is working like SQL for relational database, this paper using SPARQL to convert unstructured RDF data to traditional dataset. The above sample is tested on (Figure 6) RDF dataset which has the results in the following table.

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Table 3: traditional data-set

Transaction Title	Title	Price	Types
Transaction001	toast	0.4	food
Transaction001	egg	0.3	food
Transaction001	sausage	0.7	food
Transaction001	beans	0.5	food
Transaction001	milk	0.6	hot-drink
Transaction002	toast	0.4	food
Transaction002	egg	0.3	food
Transaction002	tomato	0.1	food
Transaction002	Juice	0.5	cold-drink
Transaction003	bread	0.3	food
Transaction003	Yogurt	0.6	food
Transaction003	tea	0.4	hot-drink
Transaction003	egg	0.3	food
Transaction003	milk	0.6	hot-drink
Transaction004	mushroom	0.5	food
Transaction004	toast	0.4	food
Transaction004	susage	0.7	food
Transaction004	tomato	0.1	food
Transaction004	juice	0.5	cold-drink
Transaction004	milk	0.6	hot-drink
Transaction005	toast	0.4	food
Transaction005	egg	0.3	food
Transaction005	beans	0.5	food
Transaction005	juice	0.5	cold-drink
Transaction005	milk	0.6	hot-drink
Transaction006	sausage	0.7	food
Transaction006	bread	0.3	food
Transaction006	mushroom	0.5	food
Transaction006	tea	0.4	hot-drink

6-3. Data Mining with Éclat algorithm

in the Data mining part, Éclat algorithm is going to apply, Éclat is one of the most famous algorithm in association rule mining. In this step, there is a traditional dataset, which is produced from RDF

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

linked data. sometimes, in this step, it needs to preprocess dataset, preprocessing data is cleaning dataset from the noisy data and arranging dataset. Finally, it needs to apply the algorithm on the dataset, then we have the result. The reason of applying Éclat algorithm in this paper is that it allows to change the format of dataset from horizontally to vertically. Also, it is not difficult to understand the result of this algorithm.

Table 4: Transaction of dataset

Transaction Title	Title	Price	Types
Transaction001	toast	0.4	food
	egg	0.3	food
	sausage	0.7	food
	beans	0.5	food
	milk	0.6	hot-drink
Transaction002	toast	0.4	food
	egg	0.3	food
	tomato	0.1	food
	Juice	0.5	cold-drink
Transaction003	bread	0.3	food
	Yogurt	0.6	food
	tea	0.4	hot-drink
	egg	0.3	food
	milk	0.6	hot-drink
Transaction004	mushroom	0.5	food
	toast	0.4	food
	sausage	0.7	food
	tomato	0.1	food
	juice	0.5	cold-drink
	milk	0.6	hot-drink
Transaction005	toast	0.4	food
	egg	0.3	food
	beans	0.5	food
	juice	0.5	cold-drink
	milk	0.6	hot-drink
Transaction006	sausage	0.7	food
	bread	0.3	food
	mushroom	0.5	food
	tea	0.4	hot-drink

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Table 5: Items inside transactions

Transaction Title	Items
Transaction001	Toast, egg, sausage, beans, milk
Transaction002	Toast, egg, tomato , Juice
Transaction003	Bread, Yogurt, tea, egg, milk
Transaction004	Mushroom, toast, sausage, tomato, juice, milk
Transaction005	Toast, egg, beans, juice, milk
Transaction006	Sausage, bread, mushroom, tea

this dataset needs to convert from horizontally to vertically, the following result is vertically of the above dataset.

Table 6: Frequent one item-set

Sausage	**Transaction001, Transaction004, Transaction006**
Bread	**Transaction003, Transaction006**
Mushroom	**Transaction004, Transaction006**
Tea	**Transaction003, Transaction006**
Toast	**Transaction001, Transaction002, Transaction004, Transaction006**
egg	**Transaction001, Transaction002, Transaction003, Transaction005**
Beans	**Transaction001, Transaction005**
Juice	**Transaction002, Transaction004, Transaction005**
Milk	**Transaction001, Transaction003, Transaction004, Transaction005**
Tomato	**Transaction002, Transaction004**
Yogurt	**Transaction003**

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Table 7: Frequent Two item-set

Sausage, Bread	**Transaction006**
Sausage, Mushroom	**Transaction004, Transaction006**
Sausage, Tea	**Transaction006**
Sausage, Toast	**Transaction001, Transaction004, Transaction006**
Sausage, Egg	**Transaction001**
Sausage, Beans	**Transaction001**
Sausage, Juice	**Transaction004**
Sausage, Milk	**Transaction001, Transaction004**
Sausage, Tomato	**Transaction004**
Bread, mushroom	**Transaction006**
Bread, tea	**Transaction003, Transaction006**
Bread, Toast	**Transaction006**
Bread, egg	**Transaction003**
Bread, milk	**Transaction003**
Mushroom, tea	**Transaction006**
Mushroom, toast	**Transaction004, Transaction006**
Mushroom, juice	**Transaction004**
Mushroom, milk	**Transaction004**
Mushroom, tomato	**Transaction004**
Tea, toast	**Transaction006**
Tea, egg	**Transaction003**
Tea, milk	**Transaction003**
Toast, egg	**Transaction001, Transaction002**
Toast, beans	**Transaction001**
Toast, juice	**Transaction002**
Toast, milk	**Transaction001, Transaction004**
Toast, tomato	**Transaction002, Transaction004**
Egg, beans	**Transaction001, Transaction005**
Egg, juice	**Transaction002, Transaction005**
Egg, milk	**Transaction001, Transaction003, Transaction005**
Egg, tomato	**Transaction002**
Beans, Juice	**Transaction005**
Beans, milk	**Transaction001, Transaction005**
Juice, milk	**Transaction004, Transaction005**
Juice, tomato	**Transaction002, Transaction004**
Milk, tomato	**Transaction004**

We will remove all pear items that only have one support.

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Table 8: Frequent two item-set with removing one support items

Sausage, Mushroom	**Transaction004, Transaction006**
Sausage, Toast	**Transaction001, Transaction004, Transaction006**
Sausage, Milk	**Transaction001, Transaction004**
Bread, tea	**Transaction003, Transaction006**
Mushroom, toast	**Transaction004, Transaction006**
Toast, egg	**Transaction001, Transaction002**
Toast, milk	**Transaction001, Transaction004**
Toast, tomato	**Transaction002, Transaction004**
Egg, beans	**Transaction001, Transaction005**
Egg, juice	**Transaction002, Transaction005**
Egg, milk	**Transaction001, Transaction003, Transaction005**
Beans, milk	**Transaction001, Transaction005**
Juice, milk	**Transaction004, Transaction005**
Juice, tomato	**Transaction002, Transaction004**

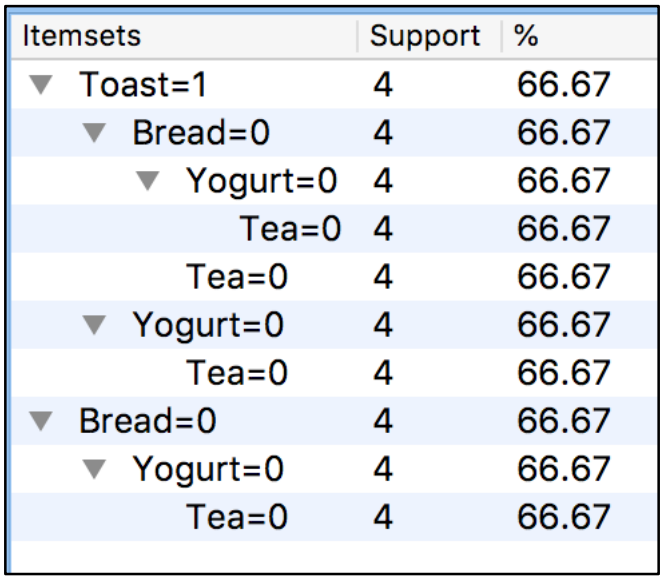
Finally, we will have three items together a shows in the followings,

Table 9: Frequent Three item-set

Sausage, Mushroom, Toast	**Transaction004, Transaction006**
Sausage, Toast, Milk	**Transaction001, Transaction004**
Sausage, Toast, Mushroom	**Transaction004, Transaction006**
Toast, tomato, Juice	**Transaction002, Transaction004**
Egg, beans, Milk	**Transaction001, Transaction005**

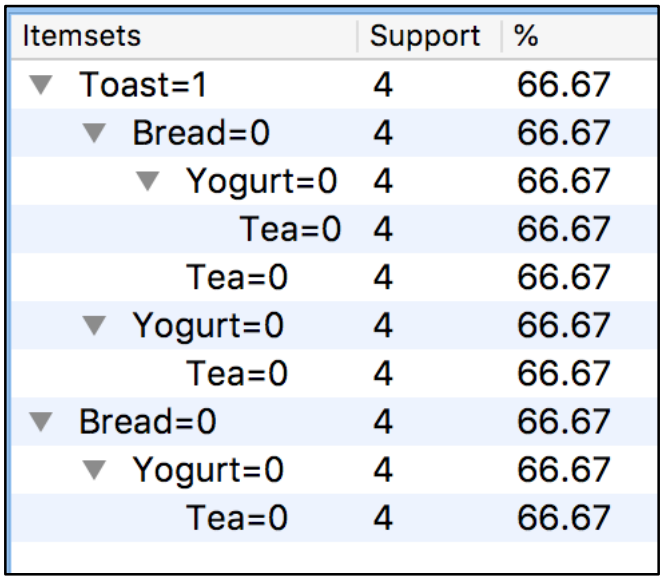
In the first item frequency, the number of item in each transaction was counted which is for three reputations for Sausage that available in (**Transaction001, Transaction004, Transaction006**), also, bread, mushroom, tea and beans only available in two Transactions. It means, support for these items are two out of 6 transactions. Furthermore, Toast, juice and sausage have three reputations in transactions, it means each of them has 50% support. As well as, Milk has four reputations, it has 66.6% support. It can clearly see that two item-sets frequency can have different repetition, for example sausage with toast has three support out of six. Toast with milk, toast with tomato, egg with beans, egg with juice, beans with milk, juice with milk and juice with tomato have two item-sets frequency. Moreover, in different dataset is available which has different support. For instance, three different example will show in the Figure 8 and Figure 9.

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)



Itemsets	Support	%
Toast=1	4	66.67
Bread=0	4	66.67
Yogurt=0	4	66.67
Tea=0	4	66.67
Tea=0	4	66.67
Yogurt=0	4	66.67
Tea=0	4	66.67
Bread=0	4	66.67
Yogurt=0	4	66.67
Tea=0	4	66.67

Figure 8: Support rate for products



Itemsets	Support	%
Toast=1	4	66.67
Bread=0	4	66.67
Yogurt=0	4	66.67
Tea=0	4	66.67
Tea=0	4	66.67
Yogurt=0	4	66.67
Tea=0	4	66.67
Bread=0	4	66.67
Yogurt=0	4	66.67
Tea=0	4	66.67

Figure 9: support rate for Sale records

Moreover, the rate of support for each items can be clearly seen in the following Figures, Figure 10 shows transaction 1 support for each items which have five products, but both Figure 11, Figure 15 include transaction002 with transaction006 respectively have four items. Also, Transaction004 in Figure 13 shows six items. Also, Figure 12, Figure 14 show five items for each of them.

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

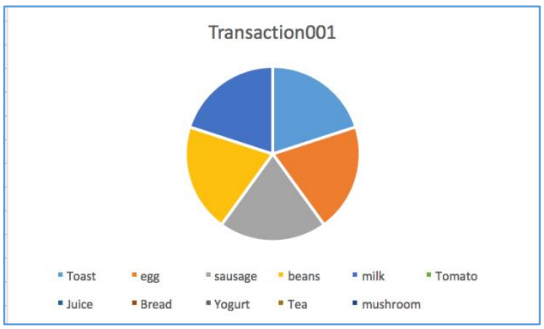


Figure 10: Transaction001 the support of item-set

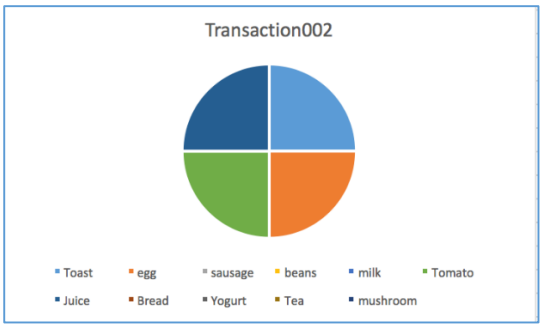


Figure 11: Transaction002 the support of item-set

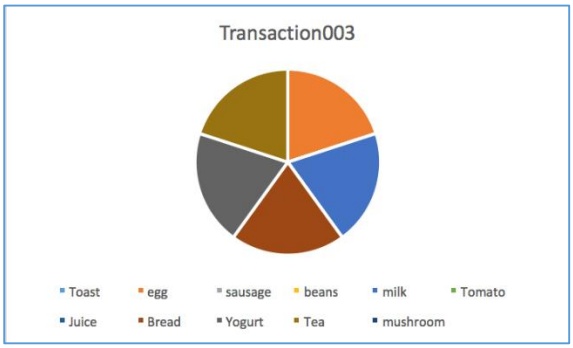


Figure 12: Transaction003 the support of item-set

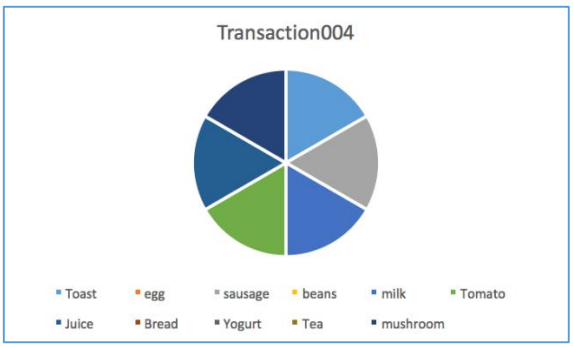


Figure 13: Transaction004 the support of item-set

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

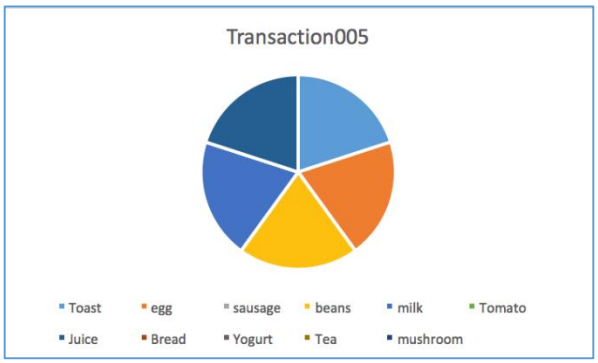


Figure 14: Transaction005 the support of item-set

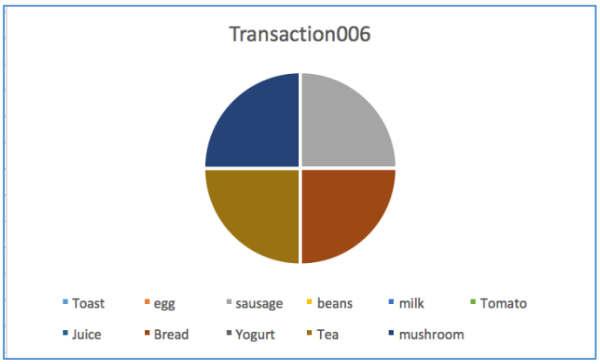


Figure 15: Transaction006 the support of item-set

Finally, it shows that {Sausage, mushroom, toast} has a support of 2 because it shows in transactions Transaction004 (T4) and Transaction006 (T6). Moreover, it can be seen that Egg with milk has three support because it appears in Transaction001 (T1), Transaction003 (T3) and Transaction005 (T5) as shows in Figure 16.

DOI: http://dx.doi.org/10.25098/3.2.21

The Scientific Journal of Cihan University – Slemani **PP: 01-25**
Volume (3), Issue (2), December 2019
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

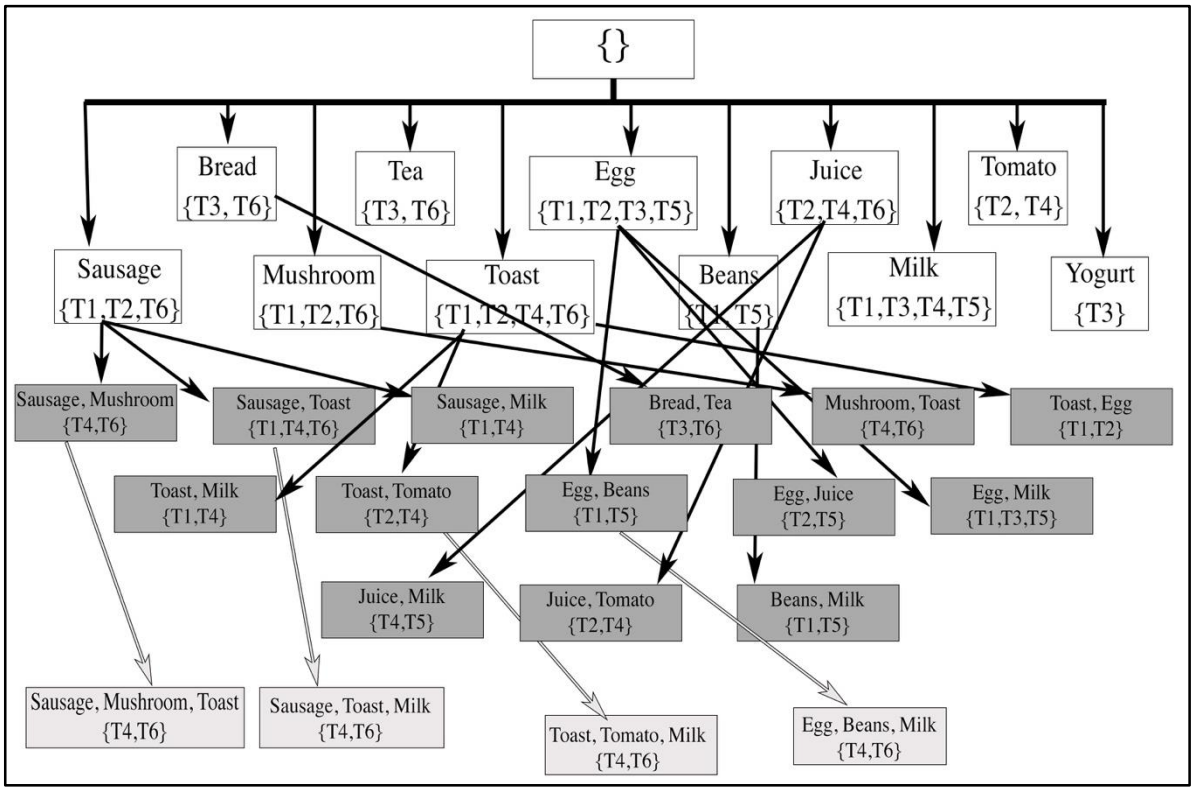


Figure 16: Tree of Éclat algorithm

Discussion:

This paper detects the impact of mixing both RDF data with Éclat algorithms which are parts of semantic web and data mining respectively. At the beginning of the process, the RDF data must check for validity, and then the data converted into traditional data using SPARQL query language. As a result, the pure traditional dataset produced, because data preprocessing applied after obtaining the traditional dataset. Secondly, Éclat algorithm is discovering frequent item sets. Éclat is going to change the dataset from horizontally to vertically before mining. Item in Éclat algorithm has support without confidents, support is how many times the item set shows in a transaction. The RDF structures consist of three different components (subject, predicate and object) which show in figure 1. Next, RDF needs SPARQL query language to convert RDF into traditional dataset, it shows in figure 2. Éclat is an algorithm of Market basket analysis to detect frequent item-set, the support of item-set is available but it does not need confident, the item-set frequent is shown in figure 4. In the Éclat algorithm first it needs to know about the number of items inside any transaction a shows in table 1, and then find the support of any item as shows in table 2. Then it needs to find the two item-set and then it needs to increase item-set as necessary as shows in table 6,7 and 8.

Conclusion:

In conclusion, combining both area semantic web and data mining has an interesting result, the techniques of this paper is mining RDF data, and the data mining algorithm is Éclat algorithm because it is frequently item-set. Éclat is one of the famous algorithms for association rule mining. Also, Éclat is unsupervised learning, it can find the numbers of item inside any transaction, which find the support of any item-set. RDF data, which has (subject, predicate and object). Also, applied Éclat algorithm is needs traditional dataset which is produced from RDF data using SPARQL. Also, Éclat is part of Association rule mining. In this paper, firstly it needs RDF data then it has to be changed into traditional dataset, the conversion of this dataset needs SPARQL which is query language. Then the process needs to apply Éclat algorithm, finally the result of mining RDF data using Éclat has been produced. This combination is significant to mining semi-structured dataset such as RDF data. However, several questions still need to be solved; the issue of big data with data mining and semantic web is an intriguing one which could be usefully explored in further research.

Reference:

1. AbedjanEmail, Z., & Naumann, F. (2013). Improving RDF Data Through Association Rule Mining. *Datenbank Spektrum , 13* (2).
2. Aggarwal, C. C., & Han, J. (2014). *Frequent Pattern Mining.* Springer.
3. Baratia, M., Baia, Q., & Liub, Q. (2017). Mining semantic association rules from RDF data . *133*, 183-196 .
4. Borgelt, C. (2012). Frequent item set mining . *WIREs Data Mining And Knowledge discovery , 2* (6), 437–456.
5. DECKER, S., MELNIK, S., HARMELEN, F. V., FENSEL, D., KLEIN, A. M., BROEKSTRA, J., et al. (2000). The Semantic Web: The Roles of XML and RDF. *IEEE Internet and Computing , 4* (5), 63-74.
6. DuCharme, B. (2013). *Learning SPARQL: Querying and Updating with SPARQL 1.1* (Second edition ed.). O'Reilly Media.
7. Gayo, o. E., Prud'hommeaux, E., Boneva, I., & Kontokostas, D. (2018). *Validating RDF Data* (Vol. 7). Morgan & Claypool.
8. Han, J., Pei, J., & Kamber, M. (2012). *Data Mining: Concepts and Techniques* . Waltham, MA: Elsevier Inc.
9. Hidber, C. (1999). Online Association Rule Mining . *Proceeding SIGMOD '99 Proceedings of the 1999 ACM SIGMOD international conference on Management of data , 28* (2), 145-156 .
10. kaur, M., & Grag, U. (2014). ECLAT Algorithm for Frequent Item sets Generation. *International Journal of Computer Systems , 1* (2).
11. Koh, Y. S., & Rountree, N. (2010). *Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection* . SCOPUS.
12. Maimon, O., & Rokach, L. (2005). *Data Mining and Knowledge Discovery Handbook* . springer.

DOI: http://dx.doi.org/10.25098/3.2.21

13. Mohammed, W. M., & Saraee, M. M. (2016). Mining Semantic Web Data Using K-means Clustering Algorithm. *British Journal of Mathematics & Computer Science , 13* (1), 1-14.
14. Mohammed, W., & Saraee, M. M. (2016). Sematic Web Mining Using Fuzzy C-means Algorithm. *16* (4), 1-16.
15. ReynaudEmail, J., Toussaint, Y., & Napoli, A. (2019). Using Redescriptions and Formal Concept Analysis for Mining Definitions in Linked Data. *International Conference on Formal Concept Analysis , 11511*, 241-256.
16. Shimada, K., Hirasawa, K., & Hu, J. (2006). Association Rule Mining with Chi-Squared Test Using Alternate Genetic Network Programming . *Advances in Data Mining. Applications in Medicine, Web Mining, Marketing, Image and Signal Mining* , 202-216.
17. Venkata, N. p., Kappara, P., & Ichise, R. (2011). LiDDM: A Data Mining System for Linked Data . *Workshop on Linked Data on the Web , 813*.