ISSN: 2350-0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 5, Issue 8 , August 2018

Mapping the Intellectual Structure of the Linked Data Field: A Co-Word Analysis and Social Network Analysis

Aint Thin Zar Kyaw, Zhongyi Wang

Tutor, Department of Library and Information Studies, University of Yangon, Yangon, Myanmar
Master Student, School of Information Management, Central China Normal University, Wuhan, Hubei, China
Associate Professor, School of Information Management, Central China Normal University, Wuhan, Hubei, China

ABSTRACT: The study focuses on the intellectual structure of linked data field by using co-word analysis and social network analysis. For this research, the data were retrieved from Web of Science (WOS) core collection database of linked data of Web of Science, including 946 articles with 2332 keywords. Co-word analysis utilized co-occurrence matrix for calculating factor analysis with principal component analysis, cluster analysis by Ward's method, similarity of proximity matrix of Pearson Correlation Coefficient and multidimensional scaling (MDS), PROXCAL algorithm with the aids of SPSS 23. Moreover, through co-occurrence matrix, social network analysis discovered the measure of centrality and k-core analysis. Therefore, the paper discovers that co-word analysis can investigate subject clusters in linked data, and can compare between subjects and intellectual structure by dividing into nine clusters which are "Internet of things", Entity linking", "Education", "Semantic Web", "linked data", "Web of data", "Dbpedia", "Data integration" and "Ontology". According to the results, both analysis can make and understand more about linked data field and how those fields linked each other with the subject. Briefly, this research improves the relative effectiveness development of using linked data in different years of articles and it aims to know how the keywords focus in the linked data field.

KEYWORDS*:* Linked Data, Co-word Analysis, Social Network Analysis, K-core Analysis

I. INTRODUCTION

Linked data is an approach of using URI, RDF, SPARQL and semantic web to publishing and sharing data on the Web (Dave Reynolds, 2016). Linked data has the ability to publish and connect the structured data on the Web for supporting a global information space of linked documents to one where both documents and data are linked (Christian Bizer, et al. 2009).

Many researchers have been done to map intellectual structure of different subject fields in different ways. Going through literature or internet, linked data fields seems not to be mapped in any ways. Then, there are many different kinds of articles which are concerned with linked data field. There are also many subcategories and a number of widely-used papers which are in connection among themselves. However, the distribution of different subject ideas in the domain or the related subjects may not be explained yet. Therefore, this study searches mapping intellectual structure of linked data field by using co-word analysis and social network analysis. Mapping can make this field more under stable on how those fields linked with each other within the subjects. Co-word analysis can be employed in identifying the discovered domain of knowledge quantitatively and the relations between domains. On the other hand, social network analysis measures the flows of relationships of each keyword. So, this paper examined linked data related research areas and trends through co-word analysis and social network analysis. In this communication, intellectual structure of linked data is defined as supporting a cluster through Cluster Analysis with Pearson's Correlation coefficient, Multidimensional scaling (MDS), measure of centrality and k-core analysis which are based on co-occurrence matrix with the aids of SPSS 23 and UCINET. It hopes that scholars and students will understand and have a practical understanding of the direction subjects in linked data field.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

II. RELATED WORK

Linked Data

Linked Data is a method of publishing structured data as it is essential and interconnects through semantic queries. Linked data is created by HTTP, RDF, SPARQL and URIs that can make between data from different sources automatically by means of computers. The linked data depends on a number of key web standards such as URIs, RDF, SPARQL, Semantic Web, etc. Linked data can help in the communication and distribution of structured data on the Web.

A. Linked Data Principles

Linked data principles can be said to be a set of best practices for publishing and connecting structured data on the Web. According to linked data principles, users can generally involve in publishing a data set and interlinking it with existing data sets.
Tim Berners-Lee outlined four principles of linked data in his "Linked Data" note of 2006, paraphrased along the following lines:
1) Use URIs to name (identify) things.
URI (Uniform Resource Identifier) is unique identification for all things. If we have the same URI only, it does not confuse and also incorporate in various dataset.
2) Use HTTP URIs so that these things can be looked up (interpreted, "dereference").
3) Provide useful information about what a name identifies when it's looked up, using open standards such as RDF, SPARQL, etc.
RDF is one part of WORLD WIDE WEB Consortium (W3C), which is generally used in conceptual description or modelling of information.
4) Refer to other things using their HTTP URI-based names when publishing data on the Web.

B. Publishing Linked Data on the Web

Actually, linked data principles are for publishing data on the Web. There are also three basic steps that involves the publishing data set as linked data on the Web as following: (Christian Bizer et.al, 2009).
1) Assign URIs to the entities described by the data set and provide for dereferencing these URIs over the HTTP protocol into RDF representations.
2) Set RDF links to other data sources on the Web, so that clients can navigate the Web of Data as a whole by following RDF links.
3) Provide metadata about published data, so that clients can assess that quality of published data and choose between different means of access.

C. Linked Data Technologies

Linked data is arranged on two technologies which are Uniform Resource Identifiers (URIs) (Berner-Lee et al., 2005) and the Hypertext Transfer Protocol (HTTP) (Fielding et al.1999). A technology is critical to the Web of data which increased URIs and HTTP. The HTTP protocol is retrieval mechanism and the RDF data model represent resource description by identifying HTTP and URIs resources (Jacob & Walsh, 2004). URIs is a kind of address that assists to define in the world which is using http://scheme. In this way, HTTP provides retrieving resources. Therefore, a technology that is necessary to the Web of data complements URIs and HTTP (Bizer, C. et al.,).

Co-word Analysis

The co-word analysis was improved between the Centre de Sociologie de l'Innovation of the Eole National Superieure des Mines of Paris and Centre National de la Recherche Scientificque (CNRS) of France during the 1980s and was known as "LEXIMAPPE". Co-word analysis is figuring the co-occurrence frequency. This analysis identifies relationships and interactions different parts of articles. It is also a method which can be integrated the structure of specific subject fields without classification system. Otherwise, it is also one kind of technique that explores the concept network in different fields. It is useful in different scientific domains which are computer science, information science library science, business economics, mathematical computational biology and engineering, etc. If two keywords are happening at the same time in same articles, it means that two articles are related each other. Co-word analysis can be said that it is as similar as co-citation analysis and co-author analysis. Therefore, as above mentioned, this research uses co-word analysis because of the ability to analyse the intellectual structure of subject field and it collects keywords from the articles of linked data field. Then, it discovered the high frequency of keywords and co-occurrence matrix. And then, with the aid of SPSS 23, this research observes factor analysis, proximity matrix: finding similarity with Pearson's Correlation Coefficient, clustering technique by using Ward's method and multidimensional scaling (MDS) which described the structure of clusters. Moreover, co-word analysis also describes co-occurrence for pairing words that include in research. If two keywords are happening at the same time in same articles, it means that two articles are related each other. Thus, this paper aims to analysis in mapping the linked data and observe the intellectual structure of it.

Social Network Analysis

Base on Rob Cross (2006), there are three analyses which are planning and administering a network analysis, visual analysis of social network and quantitative analysis of social networks. That three are widely used in social network analysis. Among that the three, visualization provides a relation network of structure and relationships of keywords is generated by NetDraw. It concerns with trends, cluster and patterns and provides a frame of reference and a temporary storage area. In this study, k-core analysis is also a kind of visualization (JengolBeck, 2013). Then, quantitative analysis of organizational networks of measure of centrality is important which has been made available in UCINET6 (Borgatti, et al., 2005). The finally step, in social network analysis (SNA), co-occurrence matrix is chosen for using in UCINET6 and two methods are identified in this research, which are measure of centrality and k-core analysis. Social network analysis methods provide some useful tools not only for addressing one of the most important but also one of the most complex and difficult (Robert A. Hanneman, 2008). The method can express the structure of actors of low level or high level of ties and k-core analysis can investigate the tighter relationship among the members' relationships.

III. RESEARCH METHODS

There are many different research methods for different kinds of research work. In this paper, it discovered about the intellectual structure of linked data field from Web of Science by using frequency analysis, co-occurrence matrix, factor analysis, cluster analysis, multidimensional scaling, measure of centrality and k-core analysis. Based on co-word analysis, similarity of proximity matrix with Pearson's Correlation Coefficient uses co-occurrence matrix on the basis of the 30 high frequency keywords. In social network analysis, two methods which are the measure of centrality and k-core analysis by UCINET are chosen. Therefore, for mapping intellectual structure of linked data fields, the figure 1 also shows the structure of research method. More about details are as follows;

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

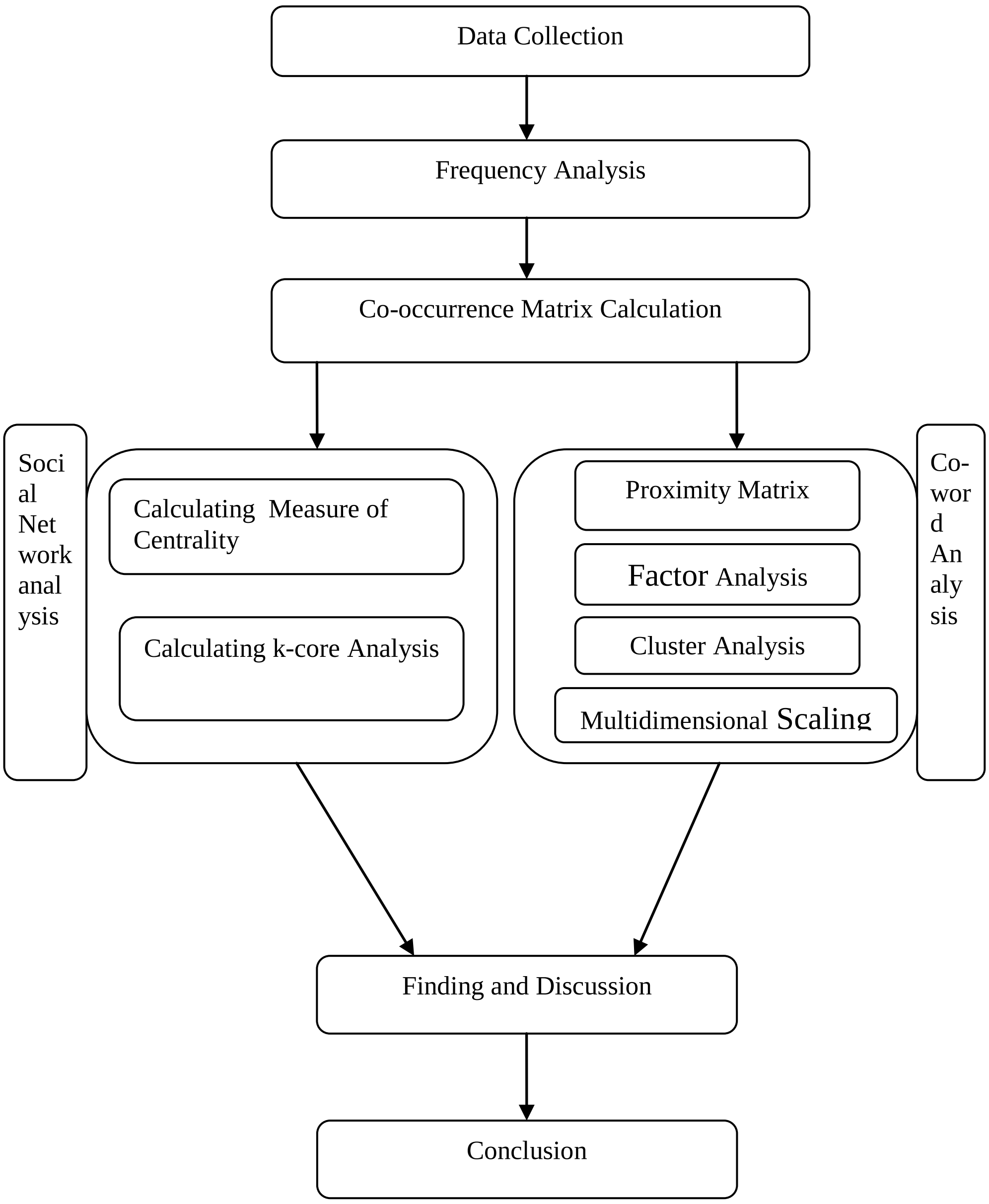

Fig1. The Process of Research Method

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

DATA RETRIEVAL STRATEGY AND KEYWORDS COLLECTION

For this research, the data were retrieved from Web of Science (WOS) core collection database of linked data of Web of Science, including 2332 keywords from 946 articles. The data were collected in December 2016.
Using those keywords, this research investigated about 30 high frequency keywords to examine the co-occurrence matrix through "Excel>Find" and counting manually by ourselves. Among 30 frequency keywords, the maximum frequency starts from 309 frequencies of "linked data" and ends at the minimum frequency of "question and answering" of 5 frequencies. Then, in Co-word analysis, it utilizes co-occurrence matrix for calculating factor analysis with principal component analysis which gives a framework for additional analysis on cluster and MDS, cluster analysis by Ward's method, similarity of proximity matrix of Pearson Correlation Coefficient and multidimensional scaling (MDS), using PROXSCAL with the aids of SPSS 23. Moreover, through co-occurrence matrix, social network analysis discovered the measure of centrality and k-core analysis with the aids of UCINET and NetDraw. Therefore, this paper discovers that co-word analysis can investigate subject clusters in linked data, and can compare between subjects and intellectual structure by dividing into nine clusters.

IV. DATA ANALYSIS AND RESULT

Frequency Analysis

According to the basic procedure of this research, the study takes the retrieved paper as explanatory data to analyze the research process of the "linked data field". The result of frequency analysis of this study is as shown in Table (1). It was calculated by Excel and it describes 30 high frequency keywords which are linked data (309), Semantic Web (125), RDF (110), Ontology (39), Data Models (32), SPARQL (23), Linked Open Data (22), Data Mining (20), data integration (12), metadata (12), and so on. This overall cognition is conscious mental activities for the main idea of linked data research.

Table1. High Frequency Keywords

No.	Frequency	Keywords
1	309	Linked Data
2	125	Semantic Web
3	110	RDF
4	39	Ontology
5	32	Data Models
6	23	SPARQL
7	22	Linked Open Data
8	20	Data Mining
9	13	Data Integration
10	12	Metadata
11	12	Open Data
12	12	Semantic Annotation
13	9	Web of Data
14	8	Data Linkage
15	8	Provenance

No.	Frequency	Keywords
16	8	Linked Data
17	8	Semantic Web
18	8	RDF
19	39	Ontology
20	32	Data Models
21	23	SPARQL
22	22	Linked Open Data
23	20	Data Mining
24	13	Data Integration
25	12	Metadata
26	12	Open Data
27	12	Semantic Annotation
28	9	Web of Data
29	8	Data Linkage
30	8	Provenance

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

Co-occurrence Matrix and Similarity of Proximity Matrix

The purpose of the co-occurrence matrix is to count the number of times that appear in each entity in rows at the same time with each entity columns. The higher co-occurrence frequency of the two keywords means a closer relationship between them (Bei-Ni-Yan, 2015). The result of the co-occurrence matrix is calculated through "Excel>Find & Select" by counting the number of times of co-occurrence of the 30 high frequency keywords are as shown in Table 2. The result shows the semantic web and linked data are the closest relationship of co-occurrence matrix in linked data field, and data linkage is non-occurrence matrix with 29 high frequency keywords. Co-occurrence matrix emphasizes on words which have relationships between them.

Table2. Part of Matrix of Co-occurring Words

	Linked data	Semantic Web	RDF	Ontology	Data Models	SPARQL	Linked Open Data	Data Mining	Data Integration	Metadata	Open Data
Linked Data	309	82	67	45	18	12	5	14	7	7	6
Semantic Web	82	125	27	21	7	5	7	5	4	3	1
RDF	67	25	110	17	16	12	7	4	2	1	4
Ontology	45	15	17	39	9	1	2	2	3	1	1
Data Models	18	7	16	9	32	0	1	2	1	1	0
SPARQL	12	9	12	1	0	23	1	0	0	0	0
Linked Open Data	5	8	6	2	1	0	22	1	0	2	0
Data Mining	14	5	4	2	2	0	1	20	1	1	0
Data Integration	7	4	4	3	1	0	0	1	13	0	0
Metadata	7	4	2	2	1	0	2	1	0	12	1

As the result of Table 3, similarity matrix can standardize the difference between keywords with high and low appearance frequency as normalizing the co-occurrence frequency range (Cho, 2014). For getting the result, this study calculates variables with Pearson's Correlation Coefficient base on co-occurrence matrix define similarity by the software IBM SPSS statistics version 23. As a result, between data management and open data of the maximum distance is 0.893 (89%) which means they have high positive correlation. The result clearly identified which keywords have high correlation and which have low correlation. In addition, the results of the numbers on the low half are the same as the number in the top half.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

Table3. Similarity of Proximity Matrix using Correlation Coefficients

	Proximity Matrix Correlation Between Vectors of Values									
	Linked Data	Semantic Web	RDF	Ontology	Data Models	SPARQL	LOD	Data Mining	Data Integration	Metadata
Linked Data	1.000	.718	.663	.833	.528	.478	.230	.586	.481	.471
Semantic Web	.718	1.000	.521	.684	.400	.398	.337	.439	.429	.380
RDF	.663	.521	1.000	.650	.626	.597	.336	.397	.337	.255
Ontology	.833	.684	.650	1.000	.613	.381	.245	.474	.491	.378
Data Models	.528	.400	.626	.613	1.000	.269	.175	.349	.291	.236
SPARQL	.478	.398	.597	.381	.269	1.000	.151	.221	.176	.136
LOD	.230	.337	.336	.245	.175	.151	1.000	.148	.066	.237
Data Mining	.586	.439	.397	.474	.349	.221	.148	1.000	.318	.310
Data Integration	.481	.429	.337	.491	.291	.176	.066	.318	1.000	.172
Metadata	.471	.380	.255	.378	.236	.136	.237	.310	.172	1.000

Factor Analysis

Factor analysis is a group of statistical techniques which can be used to analyze interrelationships among a large number of variables in terms of their common underlying factors. It also aims to decrease the number of variables and to discover the framework in the relationships between variables (Thurstone, 1931). In this result, factor analysis is managed by SPSS 23 with principal component analysis method which is related to factor analysis as shown in Table 4. The results are 31.39%, 9.835% and so on which are the percentage of variances recorded by each component. As also seen in 9^{th} rows of Cumulative Eigenvalues, it shows a value of 68.979. It means that the seven three factors together describe 68.979 of the total variance. The results accept the 30 original variables which can be chosen nine factors and the number of variables have been decreased from 30 to 9 components and it also have Eigenvalues greater than 1. Moreover, it gives a framework for additional analysis on Cluster and MDS.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

Table4. Result of the Principal Components Analysis

Total Variance Explained									
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.420	31.399	31.399	9.420	31.399	31.399	3.525	11.750	11.750
2	2.950	9.835	41.234	2.950	9.835	41.234	3.005	10.016	21.765
3	1.582	5.274	46.507	1.582	5.274	46.507	2.573	8.577	30.342
4	1.398	4.660	51.168	1.398	4.660	51.168	2.566	8.555	38.896
5	1.191	3.968	55.136	1.191	3.968	55.136	2.326	7.755	46.651
6	1.135	3.784	58.920	1.135	3.784	58.920	2.300	7.667	54.319
7	1.067	3.557	62.477	1.067	3.557	62.477	2.021	6.737	61.055
8	.992	3.306	65.783	.992	3.306	65.783	1.370	4.568	65.623
9	.959	3.196	68.979	.959	3.196	68.979	1.007	3.356	68.979
10	.927	3.090	72.069						
11	.843	2.809	74.878						
12	.747	2.490	77.368						
13	.717	2.390	79.758						
14	.703	2.344	82.102						
15	.646	2.154	84.256						
16	.609	2.030	86.285						
17	.523	1.743	88.029						
18	.470	1.566	89.594						
19	.466	1.553	91.147						
20	.436	1.453	92.600						
21	.401	1.336	93.936						
22	.380	1.268	95.204						
23	.361	1.202	96.406						
24	.310	1.033	97.439						
25	.232	.774	98.213						
26	.187	.624	98.838						
27	.173	.577	99.415						
28	.134	.446	99.861						
29	.042	.139	100.000						
30	1.227E-16	4.091E-16	100.000						

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

Extraction Method: Principal Component Analysis.

Cluster Analysis

Cluster analysis is based on data information that subscribed objects and their relationships and it is widely used in various fields such as biology, social sciences, statistics, data mining, etc. cluster analysis finds the entities within a group which is related to one another and different from unrelated entities in other groups (Pang-Ning Tan, et al. 2005). This study uses cluster analysis of hierarchical cluster is performed on the co-occurrence matrix results. As a result, it uses the Ward's method and shows with Dendogram in Figure 2.

The dendogram can be divided into big three clusters. The first cluster which forms a group includes 12 keywords from "Internet of Things" to "Semantic Search". In second cluster, there is only three keywords which are "Semantic Web", "RDF" and "Linked Data". The big third cluster which is a large group describes 15 keywords from "Data Linkage" to "Data Mining". As more details of the clusters are as shown in Table 5 which base on cluster analysis of dendogram. In each group, the table shows the result of percentage share of the occurrence frequency divided by sum of keywords. Linked data in the cluster 5 is the biggest share of 36.65%, the second is semantic web of cluster 4 with 27.64% and the third is ontology of cluster 9 with 16%. The other cluster shares which are cluster 1(2.94%), Cluster 2(2.11). Cluster 3(2.47), Cluster 6 (2.94). Cluster 7 (4.11) and Cluster 8(5.76).

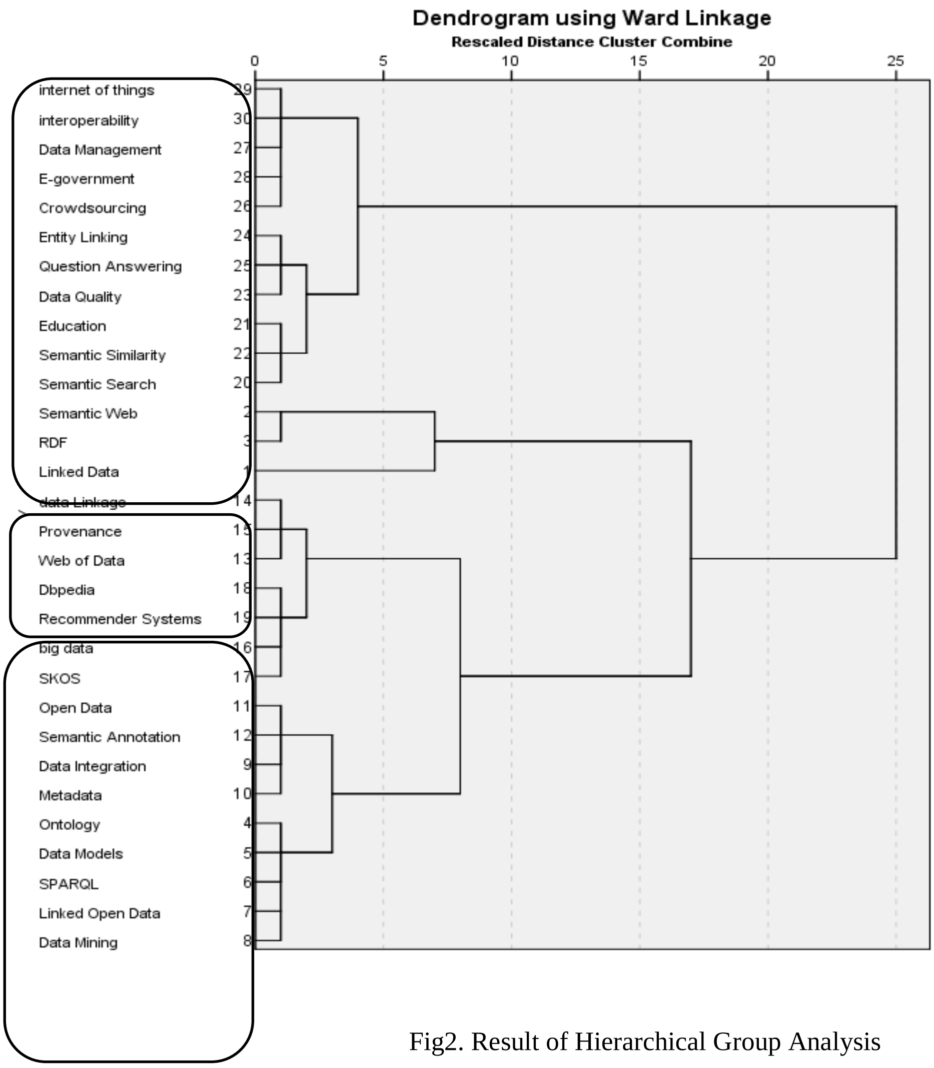

Fig2. Result of Hierarchical Group Analysis

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

Table5. Nine Clusters and Representative Keywords

Group	Cluster No.	Cluster	Representative Keywords	Share (%)
1st	C1	Internet of Things (5)	Internet of Things	2.94%
		Interoperability (5)		
		Data Management (5)		
		E-government (5)		
		Crowd sourcing (5)		
	C2	Entity Linking (6)	Entity Linking	2.11%
		Question Answering (6)		
		Data Quality (6)		
	C3	Education (7)	Education	2.47%
		Semantic Similarity (7)		
		Semantic Search (7)		
2nd	C4	Semantic Web (125)	Semantic Web	27.64%
		RDF (110)		
	C5	Linked Data (309)	Linked Data	36.65%
3rd	C6	Data Linkage (8)	Web of data	2.94%
		Provenance (8)		
		Web of Data (9)		
	C7	Dbpedia (8)	Dbpedia	4.11%
		Recommender Systems (8)		
		Big Data (8)		
		SKOS (8)		
	C8	Open Data (12)	Data Integration	5.76%
		Semantic Annotation (12)		
		Data Integration (13)		
		Metadata (12)		

Multidimensional Scaling

MDS is as the most common method (Boyack et al. 2005). It describes how data from one information system maps to data from another information system. As shown in figure 3, mapping the linked data field, multidimensional scaling map based on co-occurrence matrix with Pearson's Correlation Coefficient and it applies the PROXSCAL algorithm, and measure Euclidean distance and visualized it in two-dimensional space. The result shows the three groups among the 30 keywords. The 1st is Internet of Thing, Entity Linking and Education, the 2nd is Semantic Web, RDF and Linked Data, and 3rd group is Data Linkage, Dbpedia, Open Data and Ontology. The 2nd and 3rd big groups are together in the center. The 1st big group is widely spreading and it surrounds the other two big groups.

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

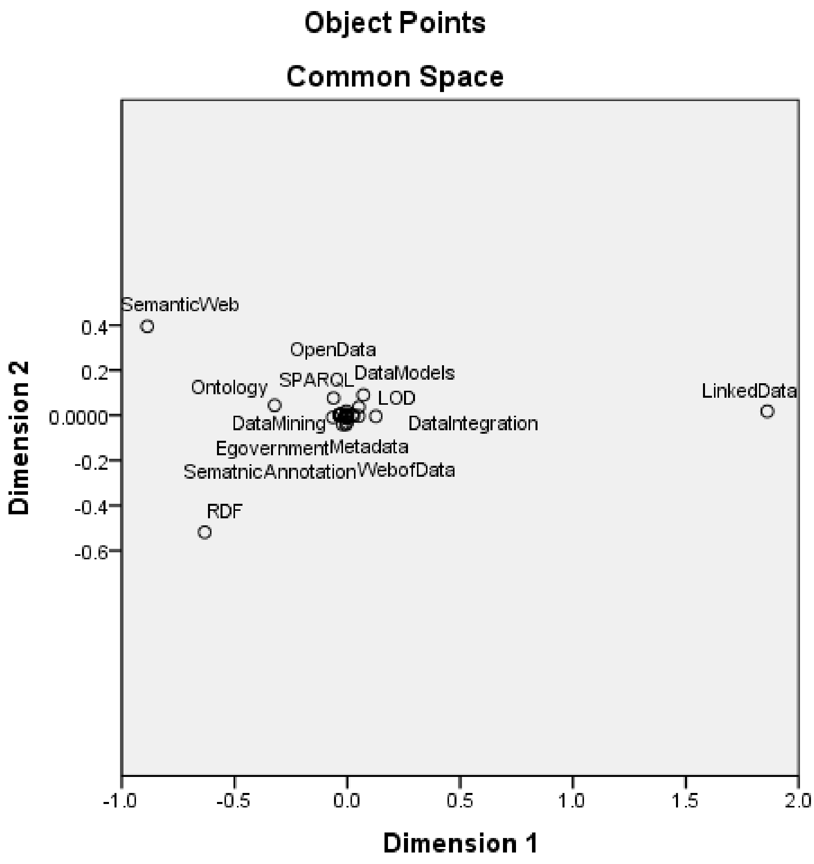

Fig3. MDS map based on keyword

Analysis of Co-Word Network

Quantitative analysis of social network analysis describes measure of centrality which are degree centrality, betweenness centrality and closeness centrality. Degree centrality shows how each node connects the network with gigantic of information. The betweenness centrality is used for finding the one which influence the flow round a system and closeness centrality (Andrew Disney, 2014). Meanwhile, k-core analysis is also used by social network analysis. In this study, co-cord correlation is analysed by UCINET6 and Net Draw to describe the measure of centrality and k-core analysis.

Table6. The 30 Keywords with Degree Centrality

ISSN: 2350-0328
International Journal of AdvancedResearch in Science, Engineering and Technology
Vol. 5, Issue 8 , August 2018

Amount	Keyword	Degree Centrality
18	RDF	748
11	Linked Data	656
29	Internet of Things	324
25	Semantic Web	323
1	Crowdsourcing	287
21	SPARQL	197
14	Ontology	154
5	Data Models	98
9	Education	92
2	Data Integration	81

Table 7: The 30 Keywords with Betweenness Centrality

Amount	Keyword	Farnesss
11	Linked Data	29
18	RDF	30
29	Internet of Things	32
25	Semantic Web	34

Table 8: The 30 Keywords with Closeness Centrality

Amount	Keyword	Betweenness Centrality
11	Linked Data	30.68
18	RDF	25.617
25	Semantic Web	18.966
29	Internet of Things	18.775
14	Ontology	13.560
1	Crowdsourcing	10.644
21	SPARQL	10.399
19	Recommender Systems	8.211
4	Data Mining	7.519
13	Metadata	5.103

ISSN: 2350-0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 5, Issue 8 , August 2018

In the result of Table 7, Table 8 and Table 9, those are all about measure of centrality. The result shows that RDF 748, linked data 656 and internet of things 324 are the highest degree centrality which are the important keywords in linked data field as a whole. In Table 8, linked data, RDF and semantic web are the highest degree of betweenness centrality. Thus, linked data, RDF and internet of things are of the highest closeness centrality as shown in Table 9. Moreover, the result of k-core analysis in Figure 4, it gets the research reveals k-core analysis network and it focuses on finding core-verge research topics. In order to display the cores clearly, nodes are marked by different colours: red square nodes(k=7) represent core themes of the network. Blue square nodes(k=6) refer to the secondary core themes. Pink square nodes (k=5) are stated between core and peripheral themes. Black square nodes (k=4) and yellow square nodes (k=3) and green square nodes (k=0) are the peripheral themes in k-core analysis. The core size is small but it interlinks more. Moreover, as shown in result, linked data node has the biggest size representing linked data which has the highest frequency of keywords, and semantic web and RDF also have higher frequency. They also have closer relationships with each other.

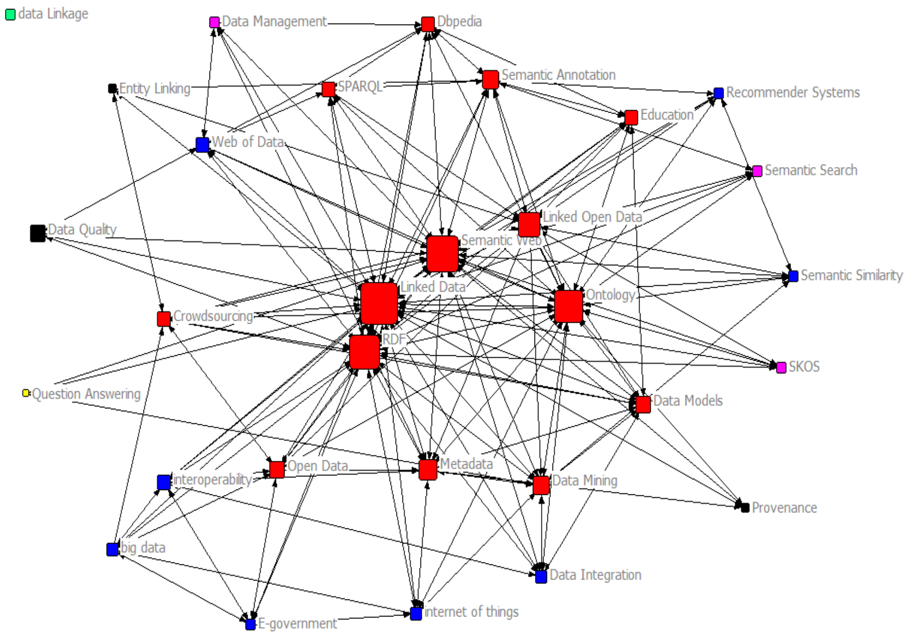

Fig4. K-core Analysis

V. FINDING AND DISCUSSION

From the above analysis, we can find that these are three big groups of linked data field as shown in Figure 2 and Table 5. Then, they are divided into nine clusters of “Internet of Things”, “Entity Linking”, ‘Education”, “Semantic Web”, “Linked Data”, “Web of Data”, “Dbpeia”, “Data Integration” and “Ontology”. Base on result of data analysis, this study describes each keywords of relationships each other and how each keyword is effective in linked data field. As the result of Table 5, in Cluster 1, the data management systems for IoT must summarize data online while providing storage, logging and adulating facilities for offline analysis. In the following Cluster 2, data quality and entity linking can be described as “Information Extraction Method” with background knowledge from the Web. There are many theoretical benefits of exploiting linked data for information extraction. Entity linking is widely used in “knowledge-based” by applying at question and answering. Knowledge base fulfils the good practices for publishing linked data. Following the principles of linked data, knowledge base corporates different sources of mobile application information that are produced and released. As shown in Cluster 3, keywords can be summarized into “Semantic Technology”.

Semantic technology uses the cloud and grid computing over linked data. As seen in Cluster 4, semantic web and RDF can be submitted as “theoretical and technological research on linked data”. Studies on this research theme mainly concentrate on “what is linked data”. Cluster 6 can be summarized as “Analysis of provenance information”. In the web content, provenance is one of the main factors which influence the trust of users. Provenance of web of data can be provided to impose qualities of linked data from the web on the application of provenance information. Analysis of provenance supplies data quality assessment of linked data and provenance relates metadata for linked data on the web.

According to Cluster 7, Dbpedia and SKOS are “Datasets of Linked Open Data” and they are based on recommender systems for using datasets. Therefore, the research theme of linked open data is linked data which is released under an open license, which does not retain its reuse for free. Open data, semantic annotation, data integration and metadata are four keywords which are gathered into Cluster 8. In Table 6 of Cluster 8, the research theme can be represented as “Information Visualization”. Information Visualization can be used as a powerful tool to facilitate the integration of open data. Metadata and data integration provide open data. Information visualization field especially concentrates on information published under the linked open data principles. In Cluster 9, keywords can be summarized as “Semantic Web Technology” that provides users in order to create data stores on the web, build vocabularies and write rules for handling data. Ontology and SPARQL helps Web of linked data by using SPARQL and it can access linked open data. (Jonathan Blaney, 2017).
Therefore, based on the result of co-word analysis, this study search that the research field of linked data point out that there are nine research themes which are Data Online”, “Information Extraction Method”, “Semantic Technology”, “What’s Linked Data”, “Linked Data”, “Analysis of Provenance Information”, “Datasets of Linked Open Data”, “Information Visualization” and “Semantic Web Technology”. In each cluster, keywords and their semantic relationships are visualized.

According to this study, linked data, semantic web and RDF are in top 3 keywords and it can be said that these three keywords are the most co-occurrence with each other in research field, as well as they also have relationships and interactions between the topic researched and emerging research trends. It also explains hierarchical cluster analysis with Dendrogram and Multidimensional Scaling (MDS), factor analysis, measure of centrality and k-core analysis know how to manage the cluster groups based on frequency. This paper distinguishes on what linked data focuses, the correlation and the recent situation in the field. From the above analysis, we can learn lots of research method with SPSS. Linked data creates World Wide Web to become a global database that we know as Web of Data. Linked data is essential to solve the problems of published data. Linked data has the ability to connect structured on the Web using World Wide Web Consortium (W3C). Therefore, linked data is not difficult to mix with other form of new data knowledge.

Briefly, this study discover which keywords have relationships with linked data fields and it also find the relationships of each keywords, and how their connections are effectively supported to each other by using co-word analysis and social network analysis. As these two analyses solve to identify the intellectual structure and development range of linked data field. Co-word analysis identifies the discovered domain of knowledge quantitatively and the relations between domains. On the other hand, social network analysis measures the flows of relationship of each keywords and visualizes the structure of relationship network by k-core analysis.

VI. CONCLUSION

Depending on this research, the next generations can have much more new research techniques and can bring innovative methods and should conduct for developing linked data field or the science development. It also aims to give the users to determine the main topic in linked data field and also find easy to study more effectively about co-occurrence keywords. On the other hand, researchers and students learn how to calculate the value of index or coefficient. Each keyword shows their co-occurrence and similarity matrix. Research can improve the relative effectiveness development of using linked data in different years of articles. Users will understand the intellectual structure and relationships between subjects in linked data fields. Another fact is that it can also find out another research based on linked data field and observe many fields in scientific development area. They will give more reasonable interpretation of intellectual structures better than before.
Future research can use further classification methods in order to investigate the research techniques and direction. In addition, library linked data might be contributed and correlated in the form of linked data. Furthermore, in the future,

ISSN: 2350-0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 5, Issue 8 , August 2018

it can use co-word analysis and social network analysis to examine on the relations between domains and intellectual by clustering or conducting k-core analysis. After studying this paper, they may find out linked data fields and might concentrate on doing another relative research by using the other journal article indexes.

REFERENCES

[1] Abu-Elkheir,M., Hayajneh, M., & Ali, NA.(2013). Data management for the interne of things: Design primitives and solution. *Sensors*, 13(11), 15582-15612.

[2] Atzori, L., Lera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, 54(15), 2787-2805.

[3] Ahsan, M., Talib, MR.& Sarwar, MU. (2016). Ensuring interoperability among hetergeneous devices thrugh IoT Middleware. *Interntional Journa of Computer Scienc and Infomation Security*, 14(4), 251-255.

[4] Tinati, R., Madaan, A., & Hall, W. (2017). The role of crowsourcing in the emerging internet-of-things. *Proceedings of the 26th International Conference on World Wid Web Companion*. (pp.1667-1672).

[5] Liu, C., He, S., & Yang H. (2017). Unsupervised joint entity linking over question answering pair with global knowdledge. *Lecture Note in Computer Science*, vol. 10565, pp. 273-286. Springer, Cham.

[6] Purnamasari, D., Banowosari, LY., & Kusaumawati, RD. (2017). Semantic similarity for search engine enhancement. *Journal of Engineering and Applied Sciences*, 12(8), 1979-1982.

[7] Noia, TD., Mirizzi, R., & Ostuni, VC. (32012). Linked open data to support content-based recommender systems. *Proceedings of the 8th International Conference on Semantic Systems* (pp.1-8).

[8] Ismail, AS., Al-Feel, HM., & Hoda, M.O. (2015). Bridging the gap for retrieving Dbpedia data. *e-Technologies and Networks for Development*, 189-194.

[9] Baker, T., Bechhofer, S., & Iesac, A. (2013). Key choices in the design of simple knowledge organization system. *Web Semantics: Science, Services and Agents on the World Wide Web*, 20, 35-49.

[10] Paulo, C., Patrik, H., & Benoit, O. (2014). Open data integration_Visualization as an asset. *3rd International Conference on Data Management Technologies and Applications*. (pp.41-47).

[11] Garcia-Godoy, MJ., Lopez-Camach, E., & Navas-Delgado, I. (2016). Re-constructing hidden-semantic data models by quering SPARQL endpoints. *Lecture Notes in Computer Science*, 9827, 405-415.

[12] d'Aquin, M., Kronberger, G., & Suarez-Figueroa, MC. (2012). Combining data mining and ontology engineering to enrich ontologies and linked data. *Proceedings of the First International Workshop on Knowledge Discovery and Data-Mining Meets Linked Open Data*. (pp.19-24).

[13] Ding, Y., Chowdhury, G>G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management*, 37(6), 817-842.

[14] Yan, BN., Lee, TS., & Le., TP. (2015). Mapping the intellectual structure of the internet of things (IoT) field (2000-2014): a co-word analysis. *Scientometrics*, 105(2), 1285-1300.

[15] Passant, A. (2010). Measuring semantic distance on linking data and using it for resources recommendations. *Linked Data Meets Artificial Intelligence*. (pp.93-98).

[16] Hu, CP., Hu, JM., & Deng, SL. (2013). A co-word analysis of library and information science in China. *Scientometrics*, 97(2), 369-382).

[17] Hartig, O., Bizer, C., & Freytag, JC. (2009). Executing SPARQL queries over the web of linked data. *International Semantic Web Conference*, 5823, 293-309.

[18] Bizer, C., Heath, T., & Berners-Lee, T. (2008). Linked data on the web. *Proceedings of the 17th International conference on World Wide Web*. (pp.1265-1266).

[19] Bizer, C. (2009). The emerging web of linked data. IEEE Intelligent Systems, 24(5), 87-92.

[20] Tummarello, G., Delbru, R., & Oren, E. (2007). Weaving the open linked data. *The Semantic Web*, 4825, 552-565.

[21] Passant, A. & Laublet, P. (2008). Meaning of a tag: A collaborative approach to Bridge the Gap between Tagging and linked data. *Proceeding of the linked data on the Web*.

[22] Hakimov, S., Oto, S.A., & Dogdu, E. (2012). Name entity recognition and disambiguation using linked data and graph-based centrality scroring. *Proceedings of the 4th International Workshop on Semantic Web Information Management*. (pp.1-7).

[23] Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, 4(2), 1-22.

[24] Khasseh, AA., Soheili, F., & Moghaddam, HS. (2017). Intellectual structure of knowledge in iMetrics: A co-word analysis. *Information Processing & Management*, 53 (3), 705-720.

[25] Li, P., Jiang, Y., Wang, J., & Yin, Z. (2017). Semantic Extension of Query for the Linked Data. International Journal on Semantic Web and Information Systems, 13(4), 109-133.

[26] Meymandpour, R. & Dovis, J.G. (2016). A semantic similarity measure for linked data: an information content-based approach. *Knowledge-Based Systems*, 109, 276-293.

[27] Barbosa, A., Bittencourt, I.I., et.al (2017). The use of software tools in linked data publication and consumption: A systematic literature review. *International Journal on Semantic Web and Information Systems*, 13(4), 68-88.

[28] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441, and 498-520.

[29] Hotelling, H. (1936). Relations between two sets of variables. *Biometrika*, 28, 321-377.

[30] Polit DF Beck CT (2012). *Nursing Research: Generating and Assessing Evidence for Nursing Practice, 9th ed*. Philadelphia, USA: Wolters Klower Health, Lippincott Williams & Wilkins.

[31] Meglen, R.R. (1991). Examining Large Databases: A Chemometric Approach Using Principal Component Analysis. *Journal of Chemometrics*. 5(3):163-179.

[32] Pearson, K (1901). One Lines and Planes of Closet Fit to Systems of Points in Space. *Philosophical Magazine*. 2(11):559-572. Doi:10.1080/14786440109462720.

ISSN: 2350-0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 5, Issue 8 , August 2018

[33] Punj, Girish & Stewart, D.W. (1983). Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*, XX,134-148.
[34] Hanneman, R.A. Intoduction to social network mehods, 2005.
[35] Marcussen, Carl (2014). Multidimensional scaling in tourism literature. *Tourism Management Perspective*, 12, 31-40.
[36] F.Cox, Trevor, A.A.Cox, Michael (2000). Multidimensional Scaling. 2nd ed., 1-19.
[37] Cambrosio, A., Limoges, C. Courtial, J. P. & Laville, F. (1993). Historical scientometrics? A study in co-word analysis. *Journal of the American Society for Information Science*, 49(13), 1206-1223.
[38] Courtial, J.P., Cahlik, T. & Callon, M. (1994). A model for social interaction between cognition and action through a key-word simulation of knowledge growth. *Scientometrics*, 31(2), 173-192.
[39] M.E.J.Newman (2008). The mathematics of networks, 2008.
[40] Marcussen, Carl (2014). Multidimensional scaling in tourism literature. *Tourism Management Perspective*, 12, 31-40.
[41] https://www.sci.unich.it/francesc/teaching/network/betweenness.html
[42] https://www.geeksforgeeds.org/betwenness-centrality-centrality-measure/
[43] https://www.theanalysisfactor.com/factor-analysis-1-introduction/
[44] www.statisticsolutions.com/cluster-analysis-2/