Journal of Theoretical and Applied Information Technology
30th November 2019. Vol.97. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EXTRACTION KEYTERM IN WORK ORDER FOR DECISION SUPPORT

[1]FARID SUKMANA, [2]FAHRUR ROZI

[1] Information Technology, University of Muhammadiyah Gresik, Indonesia

[2]Infomation Technology, STKIP PGRI Tulungagung, Indonesia

E-mail: [1]faridsukmana@umg.ac.id, [2]fahrur.rozi.maron@gmail.com

ABSTRACT

Technician is actor from industry that has job to maintenance and repair all device or infrastructure in this area. The job of technician is execution work order (WO) from user that has description like problem, symptom, root cause and solution. So WO is collection of knowledge from technician. Researcher when look in deep structure of WO has a pattern that can help technician in engineering, information technology or maintenance section getting information more efficient by using them like a google as search engine for decision making. The one of process able to use it that is extraction keyterm. This concept of method like search engine to get key in the text as main idea but apply it in WO data. Previous research about WO for decision support already done, but the process is looking for pattern of data by relating problem, symptom, root cause and solution as variable. Main thinking in this research is not looking for pattern to get best solution, but how to retrieve information of WO by searching of keyterm to cluster data based on keyterm. This study using fuzzy association rule to get weighing number and candidate cluster. The result of this study is the best of setting number for minimum support to get best cluster for decision support.

Keywords: *Work Order, Keyterm, Fuzzy, Association Rule, Stemming*

1. INTRODUCTION

Global industry in recent year already take step to industrial 4 as a time how technology be able change ability of people and can help a job more efficien, effective and quickly. The prediction in 2020, development of technology will growing faster because all sector industry using internet connection to support they job. This is phenomenon make possible all production cost, maintenance cost, and administration cost will be reduce. The one aspect that be focused in this study about how to reduce maintenance cost for repairing mechine or maintenance machine by using work order for decision support. Suddenly process of maintenance mechine, beside need tools or device as support for repairing, they need technician who has a good knowledge about infrastructure in industrial company so with this ability can be able to solve all problem more quickly and optimize production process by reducing failure mechine. This is important in industry because knowledge from technician about problem in failure mechine just owned by experience technician and not found in beginner technician, one of important sector in industrial 4 is information technology that control all system technology. Because they has abilty to repair and maintenance all equipment in networking, hardware, and software in IT sector. Thinking about in each divison of company has a part IT infrastructure such as computer, hub, switch, router and other. And another system like ERP system, payroll system in sometime can be down because failure working of server. Technician be needed to handle this job immediately. But this job can be solve quickly and believe failure system be working normally if handle by advance technician that has experienced about all device in company. But when the company only has a new technican because any change structure organization or advance technician resign from company, they need time for adaptation with all problem device in company to learn knowledge. Basically experienced and beginner technican has a same knowledge, the differences is handling process and time to get solution in a problem. 1 until 2 year is range of technician to learn about all device in industrial system. But industrial 4 require technian to handle problem more efficien time. The best company in industry, they has recording of job of technician in system that contain about knowledge some experienced technician when handle a problem. This knowledge already save in system which are called Work Order (WO). This modul is a part of Entrepise Resoure Planning

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(ERP), [1-2] ERP is system that integrate all process in industry like service maintenance, account payable, inventory management and others as a part of suppy chain for reduce production or maintenance cost. [3] Maintenance is one of process that can not separate fom efficiency cost and time, all of this need tools to get target of efficiency. Some of tools will be used in this study is how to apply WO as a part of decision making to help technican to handle problem more efficien, effective and acurate. For support this study, WO from experienced technician is needed as a part to get information like search engine by extraction text from WO and then take keyterm as a key of support to get best recommendation about problem, symptom, root cause, and solution that can be used by beginner technician. The concept of search engine is a process extraction of keyterm in document, text or data as information retrieval. [4-7] Search engine using like xml or hmtl in metadata to get information about main idea in that text, then search engine use it as retrieval information for user to get recommendation about his problem. [8] Another concept of search engine using keyterm for mechine learning to get pattern of monolingual or bilingual communication to predict information in conversation [9], and not only use in english but already find in chinese. [10-13] Extraction of keyterm is one of part text mining to get text pattern in database by identification of document key through statistic process. This need to get information more quickly and accurate by take keyterm [14] that represented summary of text. [14] A text of document sometimes has different charateristics but when look in keyterm maybe same categorized, because description of text load a same information such as theme, topic, or main idea. Based on this statement, author know any of concept of keytem can use to solve a problem in work order for technian to handle they job and this study use overal f-measure to validate keyterm.

In this study has three basic process, first process is step tokenizing and removing stop word in work order data, then stemming use lovins algorithm, after that calculate number tf-idf for every keyterm. Second process after collect all keyterm from stemming ,then looking for synonim every keyterm by using wordnet, then calculate again number tf-idf by involve synonim of keyterm in a same work order or document.Third process has two methode for getting weighing every keyterm that is combination of fuzzy logic and association rule. [15-17] This research already has been proved about applying that methode in geographic, industry and healthcare that show a good result. [18] Beside that, fuzzy association rule can apply to determine relation of keyterm and candidate of term [19] by using number of membership function from inference system in fuzzy logic where weighing number result from support of asssociation rule. This method in some process of this study expect getting result that can help maintenance for IT infrastructure. Previous research [20-21] already proved about fuzzy association rule with hypernim in wordnet to compare [22] the effect of minimum support and minimum confidence in document grouping. But this study has focus using synonim in wordnet and document that be obeserve [23-25] about work order for decision support. And for supporting this study, association rule, combination trapezoidal and triangular of fuzzy logic is needed to get weighing of term.

Some work order data in this study will be observe to result some candidate keyterm and be cluster to each document. Based on previous paragraph to get result of them, need through three process that has mention before. The best result will be use for solution in decision support and the best candidat cluster that has number of f-measure higher than another f-measure when using different minimum support ,minimum condifence and similiarity as a threshold.

2. METHOD

Figure 1 show a process on this study to get candidate cluster of keyterm. This is needed to get best information in work order data for decision support.

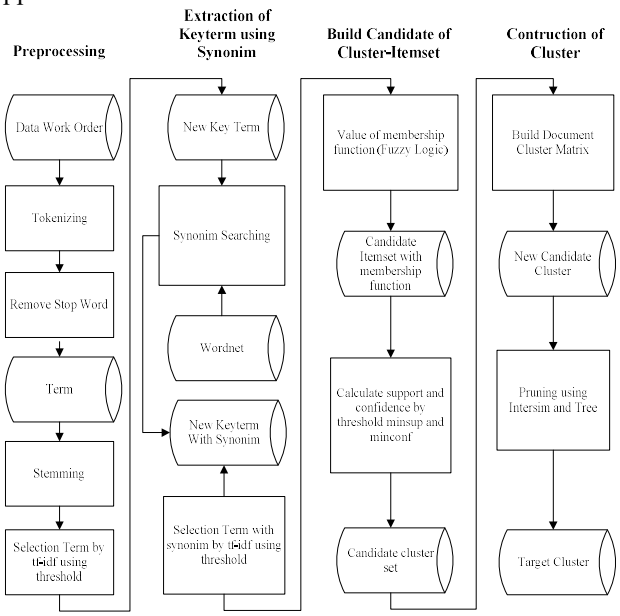

Figure. 1 : Processing of study

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.1 Tokenizing

Tokenizing is a process in sentences to remove some character that did not effect with arrangement of word and not change the meaning of sentences. The another meaning in the statement before, tokenizing is used to simplify sentences that has some token as noise in text processing so can be the easy sentences for seperating word in the next step. The mark of token that must to know such as (!@#$%^&*,:;.<>?/). Figure 2 show the tokezining process in work order.

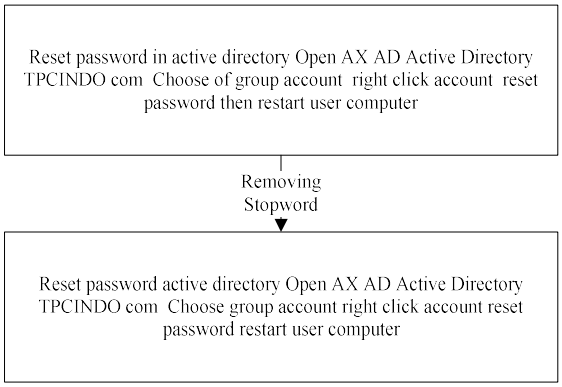

Figure 2 : Removing toke from document

2.2 Removing Stopword

Stopword is a part of preprocessing in text mining and can be ask the stopword is a noise in the sentences, because it is not a part of word, but only use as connection between word. In this research, removing stopword apply stopword classification in mysql [26]. Figure 3 show how sentence from figure 2 removing some stopword that do not need like "in", "of" and "then".

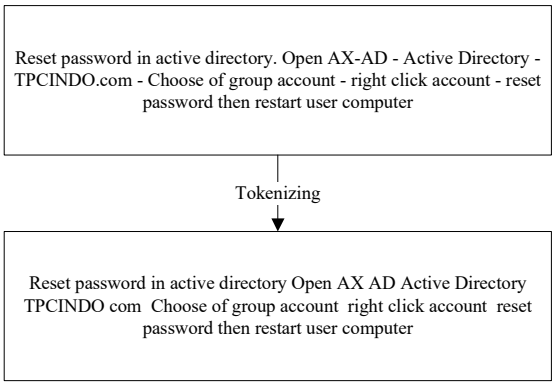

Figure 3 : Removing stopword

2.3 Stemming with Lovins Algorithm

This process is used to remove ending of word that will be result of basic of word. This research using lovins algorithm for removing ending in a word because [27] this algorithm faster than another algorithm like potter algorithm. In this algorithm only need two step to remove suffix, different with potter algorithm that need eight step to remove it. Figure 4 indicate the process of lovins algorithm in stemming process.

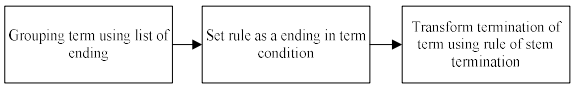

Figure 4 : Process of Lovins Algorithm for Stemming

Look at example in figure 5 about term of "Rubbing"

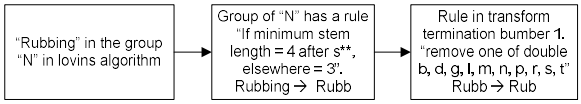

Figure 5 : Transformation "Rubbing" to "Rub"

2.4 Selection of Term

Selection of term in this research has purpose to remove some noise in some term that not too much significant to process in the next step. Because if term not often use in the documents that can be represented the term is not important. And for knowing the term important or not, any method that can use in this process. The method is using term frequency-inverse document frequency (tf-id) to calculate weighing of term. This number get from frequency of term in a document with total frequency in all documents [28].

$$tf.idf = \frac{f_{ij}}{\sum_{j=1}^{m} fij} \; x \; \log\left(\frac{|D|}{|\{di|tj \in di, di \in D\}|}\right) \text{ where } tf.idf > \propto \quad (1)$$

Where *fij* is frequency of term in a work order, then $\sum_{j=1}^{m} fij$ is frequency of term for all work order, and D is total work order in database. The result of *tf-id* is used to selection of term that has number of *tf-id* less than threshold ($\propto$). If *tf-id* not meet requirement then term be eliminate as keyterm and candidate cluster in the next step.

2.5 Wordnet

Wordnet is a database of vocabulary that content collection of word as synonym, hypernym and hyponym in lexical database [28]. The main concept of wordnet about a synset, that mean is collection of word which has the same meaning. Based on that statement this research use basic concept of wordnet as synonym of word to get some word which available in same document. Figure 5 show some synonym word for "password" in wordnet and will be search in same document about all synonym word to get a new number of tf-idf. The process explain in the next section.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1 : Synonym of "password"

Term	Synonym
password	countersign,parole,password,watchword,word

2.6 Fuzzy Association Rule

Now, fuzzy logic already growth and need to know this method has two kind of type fuzzy logic. That is fuzzy set type 1 and type 2. Fuzzy set type 1 [22] is a type of fuzzy logic that has single of membership function. Then fuzzy set type 2 is developtment from fuzzy set type 1 that has membership function more than one. [22] Fuzzy set type 2 has two concept to set membership function, that used Upper Membership Function (UMF) and Low Membership Function (LMF) when set in type of trapezoidal or triangular. But in this reseach use concept of fuzzy set type one by combine trapezoidal and triangular form. Figure 5 show about graph of triangular form of fuzzy with equation (2) and figure 6 about graph of trapezoidal form of fuzzy with equation (7). Need to know, the axis on this study use number of frequency in every term that has been resulted after calculation tf-idf every term and synonym of term.

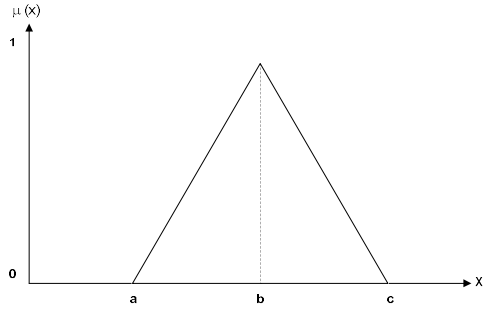

Figure 6 : Triangular

$$\mu\Delta(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases} \quad (2)$$

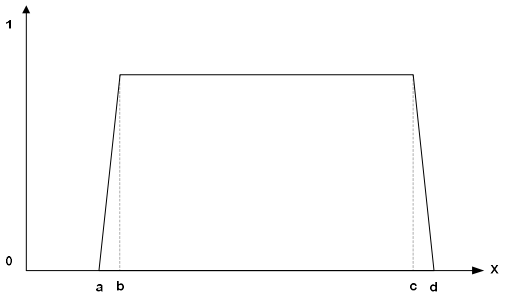

Figure 7 : Trapezoidal

$$\mu\square(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{c-x}{c-d}, & c \le x \le d \\ 0, & d \le x \end{cases} \quad (3)$$

[22] From two type graph of fuzzy logic then be combine to get maximum membership function like in equation (4). This equation use to compute support for each term. Where in the next step this number use for weighing every term.

$$W[term-j] = (\mu\Delta, \mu\square) \quad (4)$$

Term-j is term from *j*, then $\mu\Delta, \mu\square$ is number of membership function from triangular and trapezoidal form in fuzzy logic. Based on equation (4), The next step is calculate number of support every term which use membership function to get weighing number of term-j and then use it for association rule to get a number of support to be candidate cluster 1-itemset, the requirement number of support must higher than minimum support like in equation (5). Number of support obtained from total number of maximum weighing term j in each document to total of documents.

$$Support\ (term-j) = \frac{\max(\sum_{i=0}^{n} W[term-j](\mu\Delta,\mu\square))}{D} \quad (5)$$

Support must greather than minimum support. If support not meet requirement, then keyterm that be candidate cluster 1-itemset must eliminate as canidate cluster 2-itemset because number of support less than minimum support. And the keyterm that meet requirement, will process to establish as candidate cluster 2-itemset by combining keyterm from candidate cluster 1-itemset. Combination candidate cluster 1-itemset will be result candidate cluster 2-itemset, then all candidate be calculated number of support. Candidate cluster 2-itemset not be eliminated if number of support and confidence more than minimum support and confidence. Equation (6) show about getting support and confidence based on weighing from fuzzy logic.

$$sup\ (k-itemset) = \frac{\min(W[term.k],W[term.k-1],\ldots,W[term.1])}{D} \quad (6)$$

$$conf(k-itemset) = \frac{\min(W[term.k],W[term.k-1],\ldots,W[term.1])}{W[term.k]} \quad (7)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

If all process in above has been finished then candidate cluster 1-itemset and 2-itemset already generated. That is $\tilde{C}_D = \{\tilde{c}_1^1, \dots, \tilde{c}_{l-1}^2, \tilde{c}_l^q, \dots, \tilde{c}_k^q\}$, where D is document of work order, q is cluster k-itemset, then k is total cluster from candidate 1-itemset and 2-itemset.

2.7 Tree Construction

In this process involve four step, that is generate Document Term Matrix (DTM), Term Cluster Matrix (TCM), Document Cluster Matrix (DCM) and Inter_sim. All of this step be used to distinguish cluster as parent or children. [22] To calculate DTM look at equation (8).

$$W = \begin{array}{c} \\ d_1 \\ d_2 \\ \vdots \\ d_n \end{array} \begin{array}{c} \begin{array}{cccc} t_1 & t_2 & \cdots & t_p \end{array} \\ \begin{bmatrix} w_{11}^{max-R} & w_{12}^{max-} & \cdots & w_{1p}^{max-R} \\ w_{21}^{max-R} & w_{22}^{max-} & \cdots & w_{2p}^{max-R} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1}^{max-R} & w_{n2}^{max-} & \cdots & w_{np}^{max-R} \end{bmatrix} \end{array} \begin{array}{c} \\ \\ \\ \\ nxp \end{array} \quad (8)$$

Matrix (W) is maximum weighing ($[w_{ij}^{max\ \ j}]$) for every term *(t)* in each document *(d)* from membership function that resulted by fuzzy logic. And this matrix has size *nxp* that represented maximum of membership function in each document (d_n) and each term (t_p) for every cell. *n* represented of total document and p is total term which be used as keyterm. Every cell in DTM will be used to compute DCM in next step. After getting value of DTM, then continue to compute TCM using equation (9)

$$G = \begin{array}{c} \\ t_1 \\ t_2 \\ \vdots \\ t_p \end{array} \begin{array}{c} \begin{array}{cccccc} \tilde{c}_1^1 & \cdots & \tilde{c}_{l-1}^1 & \tilde{c}_l^q & \cdots & \tilde{c}_k^q \end{array} \\ \begin{bmatrix} g_{11}^{max} & \cdots & g_{1l-}^{max} & g_{1l}^{max} & \cdots & g_{1k}^{max} \\ g_{21}^{max} & \cdots & g_{2l-}^{max} & g_{2l}^{max} & \cdots & g_{2k}^{max} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ g_{p1}^{max} & \cdots & g_{pl-}^{max} & g_{pl}^{max} & \cdots & g_{pk}^{max} \end{bmatrix} \end{array} \begin{array}{c} \\ \\ \\ \\ px \end{array} \quad (9)$$

Equation 9 show about TCM that resulted from $g_{jl}^{max-R_j}$ in equation (10). $g_{jl}^{max-R_j}$ represent degree of importance level of term *tp* in candidate cluster $\tilde{c}_l^q$. This value get from comparing score value of candidate cluster with total weighing of keyterm. This matrix has size *pxk*, where p is total number of keyterm and k total candidate cluster 1-itemset and 2-itemset. And λ is number of minimum confidence.

$$g_{jl}^{max-R_j} = \frac{score\ (\tilde{c}_l^q)}{\sum_{i=1}^{n} W_{ij}^{max-R_j}},$$

where

$$score\ (\tilde{c}_l^q) = \begin{Bmatrix} \sum_{d_i \in \tilde{c}_l^1, t_j \in L_1} W_{ij}^{max-R_j}\ if\ q = 1, \\ \frac{\sum_{d_i \in \tilde{c}_l^q, t_j \in L_1} w_{ij}^{max-R_j}}{\lambda}, else \end{Bmatrix} \quad (10)$$

After value of TCM already generated then continue to calculate value of DCM. Where the value of this matrix get from multiply of DTM and TMC like in equation (11). From this equation, already get matrix of *V* as DCM that has size *nxk*, where *n* is number document or work order and *k* is number of candidate cluster k-itemset. After done to calculate DCM, continue to compute *interim* for every candidate cluster 1-itemset. This process purpose to pruning keyterm that has high similiarity in a document so can be combine to one cluster. Degree of similiarity or threshold of inter_sim set by user. If inter_sim between (c_x^1) and (c_y^1) more than threshold then merger that cluster to be one. Equation (12) show how to get *inter_sim* value.

$$inter_{sim}(c_x^1, c_y^1) = \frac{\sum_{d_i \in c_x^1, c_y^1}^{n} v_{ix} \times v_{iy}}{\sqrt{\sum_{d_i \in c_x^1}^{n} (v_{ix})^2 \times \sum_{d_1 \in c_y^1}^{n} (v_{iy})^2}} \quad (12)$$

Intersim obtained from multiply value of *v* in c_x^1, c_y^1 with square root of total square of v_{ix} and v_{iy}. Where v_{ix} and v_{iy} is value of cell in DCM.and *x* is cluster 1-itemset of first term and *y* is cluster 1-itemset of second term. And need to know value of intersim has range between [0,1].

3. ILLUSTRATION EXAMPLE

This example is a part of term in work order data which just take to use as illustration. In this example, using threshold of tf-idf is 0.8, for minimum support and minimum confidence is 0.1 and 0.4, for inter_sim has threshold is 0.8. This is some step of illustration.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1) Combine all statement in work order data. Where is in this data already available 4 description job of work order. That is a problem, symptom, root cause and solution then from all description try to combine them be one. The result look at figure 2 before step of tokenizing.

2) Remove all token in statement of work order has been descripted in figure 2.

3) Remove all stopword that is categorized as mysql stopword. Look at figure 3 that explain about removing stop word using that function.

4) Step of temming use lovins algorithm. Look at table 2 about transformation of term.

Table 2 : Transformation Of Term

Term = ***permittion***, ending = ion, rule code = Q (take 3 letter before ending and don't remove if ending after l or n), then ending is 'tt'. Using transformation rule of ending. That is 'tt' be 't'. The result is ***permit***
Term = ***make***, no ending then use as basic term. That is ***make***
Term = ***application***, ending = ication, rule code = G, ending of stem is not f so still use ***application*** as term
Term = ***password***, no ending then use as basic term. That is ***password***

5) Calculate tf-idf using equation (1) for all key term in every work order data. Table 3 example of calculate tf-idf by using 0.8 for threshold. The result true if tf-idf more than threshold.

Table 3 : Using tf-idf

WO	Term	*fij*	Σ *fij*	Σ D	di|tj∈di, di∈D	tf-idf	Result
1	permit	2	42	330	37	0.104199	True
2	make	1	26	330	32	0.089744	True
3	application	1	28	330	23	0.095129	True
5	password	3	20	330	34	0.34091	True

6) Get synonym of term by using wordnet database that explained in section 2. Look at table 4 about synonym of term table 1.

Table 4 : Synonym Of Term Using Wordnet

Term	Synonym
permit	license,permission,permit,trachinotus falcatus,licence,allow,countenance,let,tolerate
make	make,shuffle,shuffling,brand,make water,micturate,pass water,pee,pee-pee,piddle,piss,puddle,relieve oneself,spend a penny,take a leak,urinate,wee,wee-wee,ca-ca,crap,defecate,shit,stool,take a crap,take a shit,get,make up,establish,lay down,draw,cause,have,induce,stimulate,make believe,pretend,score,seduce,create,do,produce,build,construct,cook,fix,prepare,ready,give,hold,throw,arrive at,attain,gain,hit,reach,get
application	application,diligence,coating,covering,practical application,lotion,application program,applications programme
password	countersign,parole,password,watchword, word

7) Compute *tf-idf* again by involving all synonym term. That's mean number of fij include frequency of synonym term that contain in same document. For instance, "permit" has same meaning with "allow". So if find "allow" in a same document with number of term is 2, then "permit" is 3. So frequency of "permit" is 5. In table 5, the value of *tf-idf* same with table 2, because all synonym in table 3 not found in every document which used by that keyterm. All result of *tf-idf* is more than threshold so set true for all term.

Table 5: tf-idf After Using Synonym Of Wordnet

WO	Term	*fij*	Σ *fij*	Σ D	di|tj∈di, di∈D	tf-idf	Result
1	permit	2	42	330	31	0.11262	True
2	make	1	26	330	29	0.09353	True
3	application	1	28	330	22	0.09672	True
5	password	3	20	330	33	0.34538	True

8) After get keyterm from (g) then using fuzzy logic, try to find all value of membership function every keyterm in every document. Table 6 is illustration of membership function of keyterm in every document.

Table 6: Membership Function Of Keyterm

		Trapezoidal			Triangular		
WO	Term	L	M	H	L	M	H
1	permit	1	0.222	0	0.667	0	0
8	permit	1	0.222	0	0.667	0	0
1	make	0.889	0.444	0	0.333	0	0
2	make	1	0	0	1	0	0
3	application	1	0	0	1	0	0
4	application	1	0	0	1	0	0
5	password	0.889	0.444	0	0.333	0	0
6	password	0.889	0.444	0	0.333	0	0

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

9) From table 6 then get total value of membership function for every keyterm and then get maximum value of them using equation (5). This value is used to get weighing for cluster candidate 1-itemset. Table 7 indicate weighing for every keyterm.

Table 7: Weighting Of Keyterm

	Trapezoidal			Triangular			W
Term	Σμ L	Σμ M	Σμ H	Σμ L	Σμ M	Σμ H	Max (Σμ)
permit	2	0.444	0	1.333	0	0	2
make	1.889	0.444	0	1.333	0	0	1.889
application	2	0	0	2	0	0	2
password	1.778	0.889	0	0.667	0	0	1.778

10) Weighing in table 6 use for getting a value of support each keyterm. Calculation for this value based on equation (5). All keyterm be use as cluster candidate 1-itemset if it has number of support more than minimum support that already setting by user. In this instance, the minimum support is 0.1 and table 8 indicate about support all keyterm more than minimum support so all of them can be cluster candidate 1-itemset.

Table 8: Number Of Support Keyterm

Term	W / Max (Σμ)	Support
permit	2	0.25
make	1.888888889	0.236111111
application	2	0.25
password	1.777777778	0.222222222

11) From previous step, cluster candidate 1-itemset already establish. Next step is generate cluster candidate 2-itemset. This process already has equation (6) to get support and confidence to calculate that value. Before we computer all value of them, this step need to establish cluster candidate 2-itemset. This cluster is combination from all cluster in previous step. Look at table 9 about combination and the result number of support and confidence. The cluster meet requirement if support>0.1 and confidence>0.4. Based on table 8 all candidate meet requirement of them so this result will be process in next step.

Table 9: Support and Confidence For Candidate Cluster 2-Itemset

Candidate Cluster 2-itemset		Wt-k	Wt-l	Sup	Conf
permit	make	2	1.889	0.2361	1
permit	application	2	2	0.25	1
permit	password	2	1.778	0.2222	1
make	permit	1.889	2	0.2361	0.944
make	application	1.8889	2	0.2361	0.944
make	password	1.889	1.778	0.222	1
application	permit	2	2	0.25	1
application	make	2	1.889	0.2361	1
application	password	2	1.778	0.222	1
password	permit	1.778	2	0.222	0.8889
password	make	1.778	1.8889	0.222	0.94121
password	application	1.778	2	0.222	0.889

12) Now, candidate cluster 1-itemset and 2-itemset already establish. Next step is calculate Document Term Matrix (DTM) refer to equation (8). Where is all number of weighing in this matrix from step (h) by using maximum number of membership function. Table 10 show thre result DTM for each term with each document.

Table 10 : Term Cluster Matrix

	permit	make	application	password
d1	1	0.889	0	0
d2	0	1	0	0
d3	0	0	1	0
d4	0	0	1	0
d5	0	0	0	0.889
d6	0	0	0	0.889
d7	0	0	0	0
d8	1	0	0	0

13) Then build Document Term Matrix using equation (9) to get value of g and (10) for build a matrix . Table 11 show the result of DTM. Start from this step as illustrative example only use candidate cluster 2-itemset from “permit”.

Table 11 : Document Term Matrix

	permit	make	application	password	permit, make	permit, application	permit, password
permit	1	0.5	0	0	1.25	0	0
make	0.471	1	0	0	1.177	0	0
application	0	0	1	0	0	0	0
password	0	0	0	1	0	0	0

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

14) From step l and m, all component to compute Document Cluster Matrix (DCM) already available. Next step is multiply TCM and DTM to get value of DCM, based on equation (11). Table 12 indication of the result DCM.

Table 12 : Document Cluster Matrix

	permit	make	applicaton	password	permit, make	permit, application	permit, password
d1	1.418	1.389	0	0	4.046	0	0
d2	0.471	1	0	0	0	0	0
d3	0	0	1	0	0	0	0
d4	0	0	1	0	0	0	0
d5	0	0	0	0.889	0	0	0
d6	0	0	0	0.889	0	0	0
d7	0	0	0	0	0	0	0
d8	1	0.5	0	0	3	0	0

15) Based on table 11 every keyterm has number maximum of *V* in every document. Look at clearly, "permit" = {d1,d8} , "make" = {d2}, "application" = {d3,d4} and "password" = {d5,d6} already contain maximum of V in one of documents. Then using that value, compute inter_sim value every pair of keyterm based on equation (12). The result of inter_sim in table 13, and indicate if all keyterm can't merge because value of inter_sim every pair of keyterm less than 0.8 as threshold of inter_sim.

Table 13 : Inter_sim Result

No	Key Term 1	Key Term 2	Inter_sim	Result
1	permit	make	0.3731249	No Merge
2	permit	application	0	No Merge
3	permit	password	0	No Merge
4	make	application	0	No Merge
5	make	password	0	No Merge
6	application	password	0	No Merge

16) Target of cluster already find in previouse step. That is {C1(permit), C1(make), C1(application), C1(password), C2(permit,make), C2(permit,application), C2(permit,password)}. Then remove all target cluster that doesn't have cluster of parent. Table 14 show about parent and child cluster in this step.

Table 14 : Parent and child cluster

Parent	Child	Child 2	Child 3
permit	permit, make	permit, application	permit, password
make	permit, make		
application	permit, application		
password	permit, password		

17) Next step is determine all candidate cluster 2-itemset (child cluster) divided or not to cluster 1-itemset (parent cluster) using V value from step (n). If value of V child cluster lower than parent cluster, remove child cluster as candidate target of cluster. Table 15 show about comparing result of child and parent cluster.

Table 15 : Comparing result of child and parent cluster

	permit	permit,make	Divided yes/no
d1	1.419	4.04584	yes
d8	1	1.25	yes
	permit	**permit,application**	
d1	1.418	0	no
d8	1	0	no
	Permit	**permit,password**	
d1	1.418	0	no
d8	1	0	no
	make	**permit,make**	
d2	1	0	no
	application	**permit,password**	
d3	1	0	no
d4	1	0	no
	password	**permit,password**	
d5	0.889	0	no
d6	0.889	0	no

18) Figure 8 show to construct tree cluster for term to document.

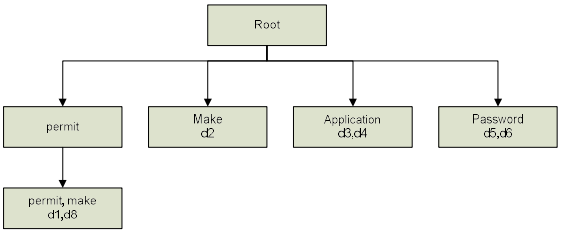

Figure : 8 Tree Cluster

4. RESULT AND DISCUSSION

This research use work order data as document to get keyterm as support decision making for technician. The modul work order already available in Microsoft Dynamic AX as ERP system in industry. This study use 720 document work order. The already save in MySQL as DBMS, and for processing method in this study already apply in PHP as programming language. The work order contain four part of decription, that is problem, symptom, root cause, and solution.

This study do not use threshold of tf-idf less than 0.6 because it need long execution time and too much use memory. Beside that not suggest

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to use minimum support more than 0.21, because number of overall f-measure is null which meaning the information of keyterm will not found in the documents. The testing in this study try to use different condition to get cluster in the document. By observing the overall f-measure in every condition. The differen condition that observe such as threshold tf-idf, minimum support, and minimum confidence. For similiarity or intersim already set 0.8 for all condition. Look at table 16 when using minimum confidence that is not effect for overall f-measure, let it use by increasing minimum confidence from left, middle and right column with different minimum support and threshold of tf-idf. When observe it, the value of overall f-measure not change. This is prove using minimum confidence is not correct decision to get best solution for extraction keyterm in this data and can ignore it for this study. Look at figure 9, observe it about using minimum support.

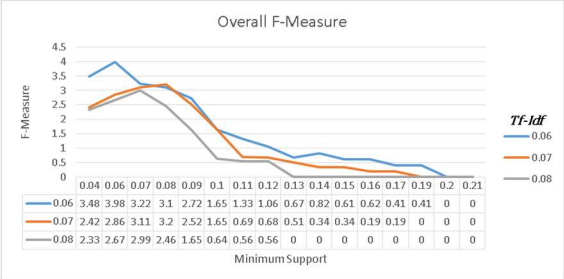

Figure 9 : F-Measure using different minimum support

Based on figure 9 about the result of overall f-measure to minimum support. When using different condition of minimum support and threshold of tf-id. That is bring the effect number of overall f-measure. The increasing minimum support make the value of f-measure going down. Look at the top number of overall f-measure by using 0.06, 0.07 and 0.08 for tf-idf . They are in position of minimum support between 0.06 until 0.08. This is proved if the best decision to get information keyterm as decision support in work order when using fuzzy association rule is using minimum support as basic to get solution.

5. CONCLUSIONS

The conclusion of the research can be summarized based on testing and analysis from the proposed method.

1) Minimum confidence is not significant to applicable in work order data using this method because not effect with overall f-measure for information retrieval.

2) Minimum support is best solution to get information about work order data because can make effect to f-measure.

3) The best solution to get information in work order is using minimum support between 0.04 until 0.12 to based on the result of overall f-measure.

This study is used to help technician to maintenance device in the manufacturer to get classification recommended solution (work order document) in the system using keyterm like the process in the google as search engine. So with this study, the technician just type the keyterm of the problem and then show the option of recommendation of solution.

REFRENCES:

[1] Butcher S.W. Assessment of Condition-Based Maintenance in the Department of Defense, Logistic Management Institute, Virginia. 2000.

[2] Zaied, R.A., Abhary K., and Gomaa A.H. *Intelegent Integrated Maintenance of Manufacturing System*. Engineering the Future. 2010, 1: 297-316.

[3] Martinez H., Laukkanen S., Mattila J. *A New Hybrid Approach for Augmented Reality Maintenance in Scientic Facilities*. International Journal of Advanced Robotic System. 2013, 10:321.

[4] E. Damiani, S. Marrara, and G. Pasi, "FuzzyXPath : using fuzzy logic an IR features to approximately query XML documents .," vol. 2, no. 1, pp. 1–11.

[5] A. Prof, S. Malallah, and Z. H. Ali, "Multi-Document Text Summarization using Fuzzy Logic and Association Rule Mining University of Technology – Department of Computer Science Al- Mustansiriya University – College of Education Department of Computer Science," no. 41, pp. 241–258, 2017.

[6] S. M. Lad, "Available Online at www.ijarcs.info Keyword Extraction from Conversation Text Document and Recommending Document using Fuzzy Logic Based Weight Matrix Method," vol. 7, no. 4, pp. 34–38, 2016.

[7] P. Lopez and L. Romary, "HUMB : Automatic Key Term Extraction from Scientific Articles in GROBID," *Proc. 5th Int. Work. Semant. Eval.*, no. July, pp. 248–251, 2010.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[8] J. Foo, Computational Terminology : Exploring Bilingual and Monolingual Term Extraction, no. 1523. 2012.

[9] J. Donghong, Y. Lingpeng, N. Yu, H. Mui, and K. Terrace, "Chinese Language IR based on Term Extraction Document Retrieval by Key Term Vectors Key Term Selection from Queries Discussion of Results and Future."

[10] S. Deepthi, "Automatic Key Term Extraction from Research Article using Hybrid Approach," vol. 166, no. 6, pp. 17–21, 2017.

[11] K. Warburton, "Glossary Creation Service leveraging state-of-the-art Term Extraction," pp. 1–5, 2013.

[12] L. Xiao, D. Wissmann, M. Brown, and S. Jablonski, "Information extraction from the web: System and techniques," pp. 195–224, 2004.

[13] A. John, "MULTI-DOCUMENT SUMMARIZATION SYSTEM : USING FUZZY LOGIC AND GENETIC ALGORITHM," vol. 7, no. 1, pp. 30–40, 2016.

[14] A. Ulanov and A. Simanovsky, "Term Validation for Vocabulary Construction and Key Term Extraction," *Proc. Int. Conf. Recent Adv. Nat. Lang. Process. 2011*, no. September, pp. 776–780, 2011.

[15] T. Bibi, Y. Gul, A. A. Rahman, and M. Riaz, "LANDSLIDE SUSCEPTIBILITY ASSESSMENT THROUGH FUZZY LOGIC INFERENCE SYSTEM (FLIS)," vol. XLII, no. October, pp. 3–5, 2016.

[16] S. C. Abou and T. Dao, "Fuzzy Logic Controller Based on Association Rules Mining : Application to Mineral Processing," vol. II, 2009.

[17] C. V Srikrishna and V. Vinod, "Fuzzy Association Rule Mining," 2015.

[18] M. J. Martin-bautista, J. M. Serrano, and M. A. Vila, "Text Mining using Fuzzy Association Rules Text Mining using Fuzzy Association Rules," no. December 2014, 2004.

[19] C. S. Series, "Fuzzy Logic Controller Based on Association Rules," vol. 37, no. 3, pp. 12–21, 2010.

[20] F. Rozi and F. Sukmana, "Metode siklis dan adaptive neuro fuzzy inference system untuk peramalan cuaca," *J. Ilm. Penelit. dan Pembelajaran Inform.*, vol. 1, no. 1, pp. 7–13, 2016.

[21] F. Rozi and F. Sukmana, "Penggunaan moving average dengan metode hybrid artificial neural network dan fuzzy inference system untuk prediksi cuaca," *J. Ilm. Penelit. dan Pembelajaran Inform.*, vol. 01, no. 02, pp. 38–42, 2016.

[22] F. Rozi, C. Fatichah, and D. Purwitasari, "Inisialisasi Klaster Menggunakan Fuzzy Association," vol. 13, no. Mcvm, pp. 190–197, 2015.

[23] F. Sukmana, and F. Rozi, "Rekomendasi Solusi Pada Sistem Computer Maintenance Management System Menggunakan Association Rule, Fisher Exact Test One Side P-Value Dan Double One Side P-Value," vol. 4, pp. 213–220, 2017.

[24] F. Sukmana and J. L. Bulaili, "Rekomendasi Solusi pada Computer Maintenance Management System Menggunakan Association Rule, Koefisien Korelasi Phi dan Chi-Square," *Pros. Semin. Nas. Manaj. Teknol. XXII*, pp. 1–8, 2015.

[25] F. Sukmana and F. Rozi, "Decision Support System on Computer Maintenance Management System Using Association Rule and Fisher Exact Test One Side P-Value," vol. 15, no. 4, pp. 1841–1851, 2017.

[26] Doyle. Damian, "MySQL Stopword", Ranks-NL, Amsterdam.

[27] A. Tordai, and Cilden E, "The Lovins Stemming Algorithm" , *The Snowball*, 2006.

[28] Chen, L. C., Tseng, F. S., & Liang, "T. Mining fuzzy frequent itemset for hierarchical document clustering. Information Processing and Management" , 193-211, 2010. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[29] Rozi F., Sukmana F, "Document Grouping by Using Meronyms and Type-2 Fuzzy Association Rule Mining", Journal of ICT Research and Applications, vol 11 hal 268-283, 2017.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$V = \begin{array}{c} \\ d_1 \\ d_2 \\ \vdots \\ d_n \end{array} \begin{bmatrix} \tilde{c}_{11}^{1} & \cdots & \tilde{c}_{1l-}^{2} & \tilde{c}_{1l}^{q} & \cdots & \tilde{c}_{1k}^{q} \\ v_{11} & \cdots & v_{1l-} & v_{1l} & \cdots & v_{lk} \\ v_{21} & \cdots & v_{2l-} & v_{2l} & \cdots & v_{2k} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nl-} & v_{nl} & \cdots & v_{nk} \end{bmatrix}_{nxk} = \begin{array}{c} \\ d_1 \\ d_2 \\ \vdots \\ d_n \end{array} \begin{bmatrix} t_1 & \cdots & t_p \\ \cdots & \cdots & \cdots \\ w_{21}^{max-} & \cdots & w_{2p}^{max-} \\ \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots \end{bmatrix}_{nxp} \cdot \begin{array}{c} \\ t_1 \\ t_2 \\ \vdots \\ t_p \end{array} \begin{bmatrix} \tilde{c}_1^1 & \tilde{c}_2^1 & \cdots & \tilde{c}_k^q \\ \cdots & g_{12}^{max-i} & \cdots & \cdots \\ \cdots & g_{22}^{max-i} & \cdots & \cdots \\ \ddots & \vdots & \ddots & \ddots \\ \cdots & g_{p2}^{max-i} & \cdots & \cdots \end{bmatrix}_{pxk}$$

(9)

Table 16 : F-Measure with different condition

Threshold tf-idf	Minsup	Similarity	F-Measure	Minsup	Similarity	F-Measure	Minsup	Similarity	F-Measure
	minconf =0.5			minconf =0.7			minconf =0.9		
0.06	0.04	0.8	3.48	0.04	0.8	3.48	0.04	0.8	3.63
0.06	0.06	0.8	3.98	0.06	0.8	3.98	0.06	0.8	3.97
0.06	0.07	0.8	3.22	0.07	0.8	3.22	0.07	0.8	3.22
0.06	0.08	0.8	3.1	0.08	0.8	3.1	0.08	0.8	3.1
0.06	0.09	0.8	2.72	0.09	0.8	2.72	0.09	0.8	2.72
0.06	0.1	0.8	1.65	0.1	0.8	1.65	0.1	0.8	1.64
0.06	0.11	0.8	1.33	0.11	0.8	1.33	0.11	0.8	1.33
0.06	0.12	0.8	1.06	0.12	0.8	1.06	0.12	0.8	1.06
0.06	0.13	0.8	0.67	0.13	0.8	0.67	0.13	0.8	0.67
0.06	0.14	0.8	0.82	0.14	0.8	0.82	0.14	0.8	0.82
0.06	0.15	0.8	0.61	0.15	0.8	0.61	0.15	0.8	0.62
0.06	0.16	0.8	0.62	0.16	0.8	0.62	0.16	0.8	0.62
0.06	0.17	0.8	0.41	0.17	0.8	0.41	0.17	0.8	0.41
0.06	0.19	0.8	0.41	0.19	0.8	0.41	0.19	0.8	0.41
0.06	0.2	0.8	0	0.2	0.8	0	0.2	0.8	0
0.06	0.21	0.8	0	0.21	0.8	0	0.21	0.8	0
Threshold tf-idf	**Minsup**	**Similarity**	**F-Measure**	**Minsup**	**Similarity**	**F-Measure**	**Minsup**	**Similarity**	**F-Measure**
	minconf =0.5			**minconf =0.7**			**minconf =0.9**		
0.07	0.04	0.8	2.42	0.04	0.8	2.42	0.04	0.8	2.42
0.07	0.06	0.8	2.86	0.06	0.8	2.86	0.06	0.8	2.86
0.07	0.07	0.8	3.11	0.07	0.8	3.11	0.07	0.8	3.11
0.07	0.08	0.8	3.2	0.08	0.8	3.2	0.08	0.8	3.2
0.07	0.09	0.8	2.52	0.09	0.8	2.52	0.09	0.8	2.52
0.07	0.1	0.8	1.65	0.1	0.8	1.65	0.1	0.8	1.65
0.07	0.11	0.8	0.69	0.11	0.8	0.69	0.11	0.8	0.69
0.07	0.12	0.8	0.68	0.12	0.8	0.68	0.12	0.8	0.68
0.07	0.13	0.8	0.51	0.13	0.8	0.51	0.13	0.8	0.51
0.07	0.14	0.8	0.34	0.14	0.8	0.34	0.14	0.8	0.34
0.07	0.15	0.8	0.34	0.15	0.8	0.34	0.15	0.8	0.34
0.07	0.16	0.8	0.19	0.16	0.8	0.19	0.16	0.8	0.19
0.07	0.17	0.8	0.19	0.17	0.8	0.19	0.17	0.8	0.19
0.07	0.19	0.8	0	0.19	0.8	0	0.19	0.8	0
0.07	0.2	0.8	0	0.2	0.8	0	0.2	0.8	0
0.07	0.21	0.8	0	0.21	0.8	0	0.21	0.8	0