PROC. INTERNAT. CONF. SCI. ENGIN.
Volume 3, April 2020 | Pages: 195-199

ISSN 2597-5250
E-ISSN 2598-232X

Information System Configuration Based On Sentence Similarity (Case Study: Islamic Boarding School ERP)

Moh. Husnus Syawab*, Melsa Khairani Nasution, Safrina Ardhillah, Muhammad Ainul Yaqin
Informatics Engineering Departement, Faculty of Science and Technology, UIN Maulana Malik Ibrahim,
Jl. Gajayana No 50 Malang 65144, Indonesia. Tel. / Fax. +62 341-551354.
Email*: husnusuus@gmail.com

Abstract. Information systems are an organized combination of people, software, hardware, communication networks and data sources in the collection, replacement, and collection of information in organizations. This configuration is required in a boarding school information system that uses SOA (Service Oriented Architecture). The configuration of the boarding school information system is based on a boarding school job description that is calculated in terms of the similarity of sentences with the business process model. Calculating the similarity of words is calculated using Word Similarity for Java (WS4J) using three semantic similarities namely wu palmer, lin, and path. From this calculation, the results of the business process model will emerge. In this study, the effect of similarity sentences in the job description of the cleanliness section is compared with the security section of the Islamic boarding school displaying that sentence similarities is almost similar, although in some job descriptions they still have a far different value for similarity

Keywords: Configuration, information systems, Islamic boarding schools, sentence similarity, service oriented architecture

Abbreviations : ERP = Enterprise Resource Planning, SOA = Service Oriented Architecture, WS4J = Word Similarity for Java

INTRODUCTION

The information system provided by Robert A. Leitch and K. Roscoe Davis (1992) is a system within an organization that helps managerial assisting operations management and meeting the needs of daily transaction processing. Besides, it is also a consideration of the strategy and evaluation activities of an organization and provides reports required by certain parties.

Islamic boarding school is one of the institutions of Islamic education held in Indonesia. Pesantren as an Islamic educational institution teaches its students (santri) to be able to understand, explore, and appreciate by discussing moral religion as a guide to community life. Pesantren has several inseparable elements, among others, Kyai as a caregiver as well as an educator, a mosque as a means of worship as well as working as a place of education, and students as students.

The configuration of the arrangement of the structure, the process of making or making the appearance of an object. Configuring information systems at Islamic boarding schools is a process to improve information systems at Islamic boarding schools so that the system becomes better.

Needed to help improve the making of appropriate information systems and following their respective fields. Constraints arise compilation of the construction of a boarding school but it does not discuss how management makes decisions and runs operations that are here. Where this system is a combination of people, information technology and procedures that have been organized so that boarding schools can be formed. This configuration explains how to design information systems that are needed in Islamic boarding schools. Most of the arrangements are needed in the world of work, not only that boarding schools also need information systems to run easily and get better progress. By making an information system design it is expected to provide a solution for the provision of Islamic boarding schools in the renewal.

In a previous study conducted by Faidul Mannan (2018) entitled the design of an accounting information system in Salafiyah Islamic Boarding School Syafi'iyah Nurul Huda Mergosono, Malang said the research objective was to discuss a good research system for pesantren and research to seek information system assistance in Islamic boarding schools.

MATERIALS AND METHODS

Study area

Data collection in this study is from secondary data. In this research the secondary data source is literature, articles, journals and websites on the internet relating to the research conducted. Secondary data sources do not directly provide data to data collectors. As secondary data from research sourced from documents consisting of notes, recorded images or photographs and observations related to the focus of this study.

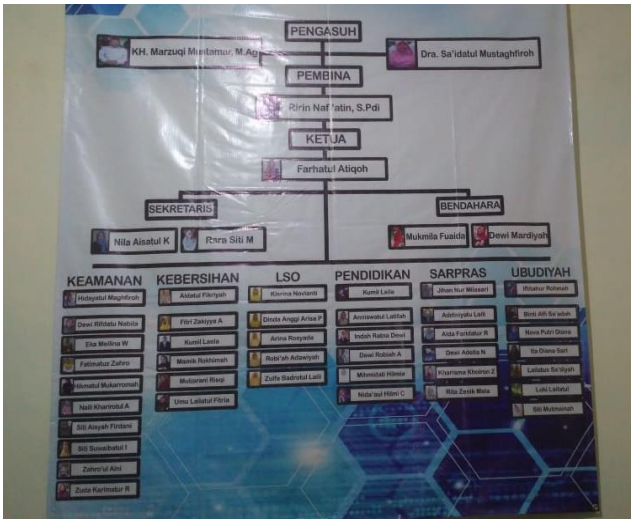

Figure 1. Organization Structure from Sabilurrosyad islamic boarding school.

Procedures

Identification organization structure

In determining the organizational structure, the organizational structure of the sabilurrosyad is used which is processed in the "Yed Graph Editor" to model the organizational structure according to their respective job description.

Identification job description

From organization structure, we used 2 Section for calculate in configure information system. The data used in this study is the job description of the Sabilurrosyad Islamic boarding school organizational structure, namely the job description from the cleanliness and security section.

Table 1. Job description from cleanliness section dan security section.

Section	Job Description
Cleanliness Section	- Schedule and allocate places that need cleaning - Manage and monitor the implementation of cleanliness and tidiness in boarding schools - Striving hard for cleanliness, coolness, neatness of the yard, and the streets of the pesantren complex, as well as coordinating with related parties including environmental sustainability
Security Section	- Monitor and control the environment, and anticipate violations, or actions that disturb public safety - Enroll students who have been real or strongly suspected of violating the rules of Islamic boarding schools

Calculating sentence similarity from section

Data processing conducted in this study is the Semantic Similarity approach which is calculated using WS4J (Word Similarity for Java) with Path-based methods namely Wu and Palmer (WUP). In this measurement, the system will calculate TF-IDF to get the lexical relevance value.

TF-IDF weighting is used to give term weight to a document. In other words, TF-IDF can also be used to find the connectedness of one term in the text. TF-IDF is represented by a vector approach, this is because each document contains a collection of words (t1, t2,..., tn). The relevance of each ti is different using the TF-IDF calculation (aris, 2015). The TF-IDF calculation combines the calculation of Term Frequency (tf) and Inverse Document Frequency (idf) as denoted in the equation:

$$W_{ij} = tf_{ij} \times id_{fj} \quad (1)$$

$$W_{ij} = tf_{ij} \times \log (D / id_{fj}) \quad (2)$$

Where W_{ij} is the term weight (tj) of the document (in). Whereas tf_{ij} is the number of occurrences of terms (tj) in the document (in). D is the number of all documents in the database and d_{fj} is the number of documents containing term (tj), which is at least one word, term (tj).

The stage will be carried out with preprocessing text. In every document in the collection, it is necessary to change the text into a matrix based on each word that has been compiled, by removing punctuation, numbers, and stopwords. Wu and Palmer are defined as similarity of two concepts based on LCS depth and shortest path. The calculation process carried out by WUP is looking for the shortest path of each concept, then each path that is formed is combined to find its lcs. Search LCS (Lowest Common Subsumer) by searching for a sense that often arises from the two paths connected. To calculate sentence similarity, the normalization process is needed by calculating the word similarity values as the formula below:

$$\text{Sentence Similarity} = (n \text{ maxx } bn) + (v \text{ maxx } bv) \quad (3)$$

'n' maxx is the average of the highest number of each column of the word similarity matrix with Noun criteria. '*v maxx'* is the average of the highest number of values in each column of the word similarity matrix with Verb criteria. Whereas for *'bn'* and *'b*v' it is the weighted value of the Noun and Verb criteria which has been calculated using AHP.

Mapping

For calculation of sentence similarity, the calculation is given, if the results of 2 sentences are 1 similarity, then the process of the business model will come out, but if the similarity is 1, the business model process will not come out. Business process models are the results of other researchers working on variations of the business process model of Islamic boarding schools.

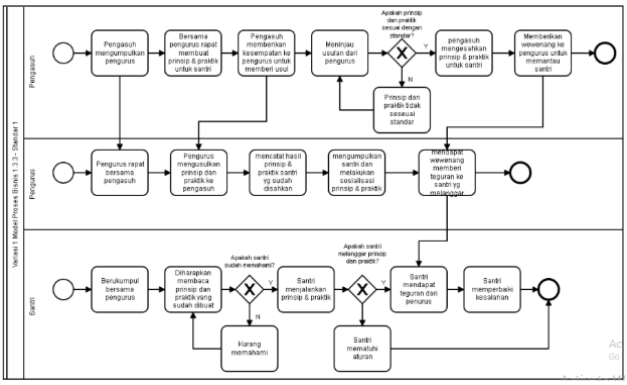

Figure 2. Business process model.

Data analysis

The data obtained is the organizational structure of the Sabilurrosyad Islamic Boarding School and job descriptions of each section. Where the organizational structure will change levels according to the job description of each section after processing is done. Other than that for sentence similarity, 2 sentences will be counted using WS4J and TF-IDF to check how much similarity between these sentences.

RESULT AND DISCUSSION

Identification organization structure

Existing organizational structure is entered into the YED graph editor and then arranged according to the job description and leaders of each section. In the YED graph editor shows the secretaries and treasurers their levels are the same as the other sections, because it shows the same leader as the others, namely the head of the cottage.

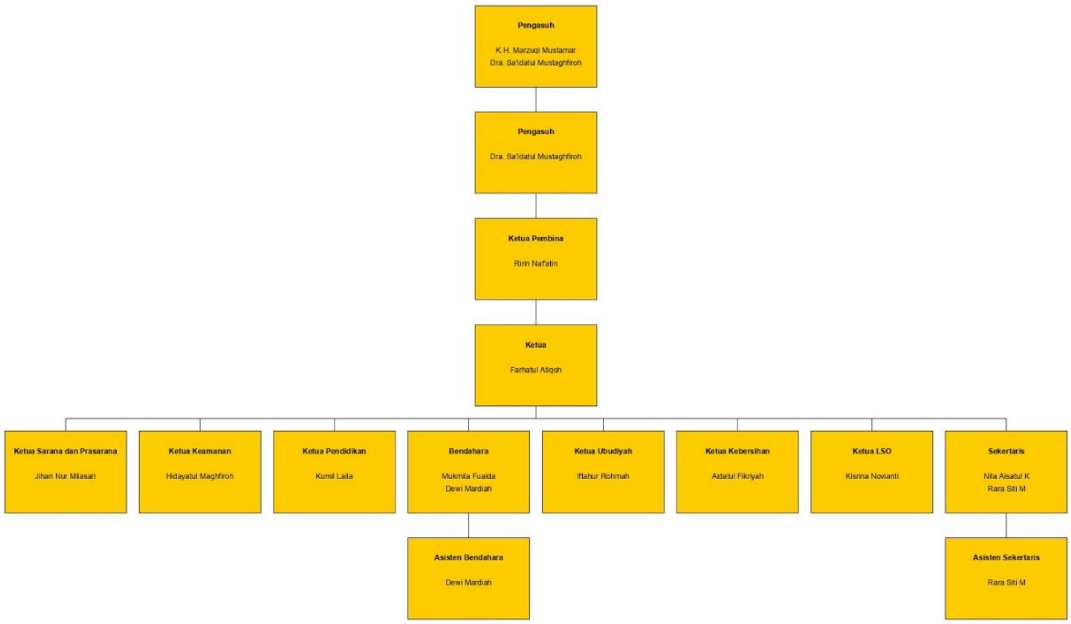

Figure 3. Organization structure YED Graph Editor.

Identification job description

The cleanliness section and the security section have their respective job descriptions. From the existing job description, one of the job descriptions taken from each section is taken, namely:

Table 2. Selected job description.

SECTION	Cleanliness Section	Security Section
JOB DESCRIPTION	Schedule and allocate places that need cleaning	Monitor and control the environment, and anticipate violations, or actions that disturb public security

Calculating sentence similarity from section

The initial stage carried out in the calculation process is the identification of sentences in the process of calculating the similarity of the sentence. This stage begins by sorting sentences that match the type or criteria of the word. The process is carried out to determine the type of words in the sentence using wordNet. In this study using WS4J (WordNet Similarity for Java) which is used to calculate the value of word similarity. In calculating the similarity of words using WS4J there is a noun class, and verbs that have similarity values calculated so that the identification phase of this sentence only chooses words in sentences based on the type of noun, and only verbs are then used

in the calculation phase of word similarity. The following is the sentence used:

- Sentence 1: Schedule and allocate places that need cleaning
- Sentence 2: Monitor and control the environment, and anticipate violations, or actions that disturb public security

The calculation of the similarity between words is done with the wu palmer approach using WS4J. The testing scenario that is done is by involving three criteria for the level of sentence similarity that is similar, moderate, and not similar. Below is an example of calculating the value of sentence similarity from the two example sentences above:

	Schedule /NNP	and /CC	allocate /DT	places /NNS	that /WDT	need /VBP	cleaning /VBG
Monitor/NNP	0.2353	-	-	0.4706	-	-	-
and/CC	-	-	-	-	-	-	-
control/VB	-	-	-	-	-	0.5000	0.4000
the/DT	-	-	-	-	-	-	-
environment/NN	0.3750	-	-	0.8750	-	-	-
,/,	-	-	-	-	-	-	-
and/CC	-	-	-	-	-	-	-
anticipate/VBP	-	-	-	-	-	0.4000	0.3333
violations/NNS	0.4444	-	-	0.7778	-	-	-
,/,	-	-	-	-	-	-	-
or/CC	-	-	-	-	-	-	-
actions/NNS	0.5000	-	-	0.8000	-	-	-
that/WDT	-	-	-	-	-	-	-
disturb/VBP	-	-	-	-	-	0.4000	0.6667
public/JJ	-	-	-	-	-	-	-
security/NN	0.5556	-	-	0.8000	-	-	-

Figure 3. Matrix of word similarity values (normalization process).

$\Sigma n = 0.5556 + 0.8750 = 1.4306$

$\overline{n} = \frac{1.4306}{2} = 0.7153$

$\Sigma v = 0.5000 + 0.6667 = 1.1667$

$\overline{v} = \frac{1.4306}{2} = 0.5833$

Sim (0.75,0.25) $= (\overline{n} \; x \; 0.75) + (\overline{v} \; x \; 0.25)$
$= (0.7153 \; x \; 0.75) + (0.5833 \; x \; 0.25)$
$= 0.6823$

Sim (0.25,0.75) $= (\overline{n} \; x \; 0.25) + (\overline{v} \; x \; 0.75)$
$= (0.7153 \; x \; 0.25) + (0.5833 \; x \; 0.75)$
$= 0.6163$

In the table above shows that the results of the sentence similarity value in (S1 - S2) is 0.6, which means the meaning of the two sentences is the same. The calculation of the sentence with the formula Sim (0.75,0.25) in the above experimental results yields 7 values higher than that of Sim (0.25,0.75), so the results show that the priority weight value of the noun is 0.75 and verb is 0.25.

Mapping

In the calculation, shows that the job description in the field of cleanliness and safety shows that the value after calculating WS4J and shows the similarity of the data after the normalization process is worth 0.6823 and 0.6163, from that value shows the difference is less than 1. This means that the 2 sentences are similar, so they will display the model process as shown in Figure 3.

The following is how to calculate the weight of the document against the query using the Term Weighting TF-IDF method using the formula described above. The following is the sentence used:

- Sentence 1: Maintain the security and cleanliness of the cottage environment
- Sentence 2: Monitor and control the environment, and anticipate violations or actions that disturb public security.

The word query (Q) used in this study is clean, security, and environment. So the number of documents in the document collection is (D) = 2.

$W_{ij} = tf_{ij} \times \log (D / d_{fi}) + 1$
$= 2 \times (\log (2/2) + 1)$
$= 2 \times (0 + 1)$
$= 2$

Thus we can get the weight value (W) for each word in the query that is in each document:

Tabel 3. Matrix term weighting TF-IDF.

	t_f		df	D/df	IDF	IDF+1	$W = t_f * (IDF + 1)$	
Q	**d1**	**d2**					**d1**	**d2**
clean	1	0	1	2	0.301	1.301	1,301	0
security	1	1	2	1	0	1	1	1
environment	1	1	2	1	0	1	1	1
							Sum (d1)	**Sum (d2)**
The weight value of each document =							3.301	2

The table above shows that the greater the value of the calculation of the weight obtained, the higher the level of similarity of documents to words (query).

CONCLUSIONS

From the results of calculations performed, shows that from the 2 sentence job description cleanliness sector and security sector have similarities with the value of 0.6823 and 0.6163 where the value is less than equal to

1, so that the business process model can be formed and become a configuration information system for boarding schools However, this is still not optimal because it has not detailed the business process model that occurs and configuration in detail, besides that, in this case, if the similarity value is more than 1, it cannot display different information system configurations, just cannot be configured.

ACKNOWLEDGMENTS

Thank you to Mr. Muhammad Ainul Yaqin as a lecturer in project management courses who have helped with this research and provided this title for research. Besides that, thank you to friends, seniors and lecturers who have also help in providing support.

REFERENCES

Abriani, Gandhis Ulta. 2018. Pengukuran Kemiripan Makna Kalimat. Islamic State University Maulana Malik Ibrahim, Malang

Informatikalogi. 2016. Pembobotan Kata atau Term Weighting IF-IDF. https://informatikalogi.com/term-weighting-tf-idf/

Lubis T. 2012. Peran Kyai dalam pengembangan Pendidikan Agama Islam di Pesantren Lirboyo Kediri. [Thesis]. Islamic State University Maulana Malik Ibrahim, Malang. [Indonesian]

Mannan F. 2018. Perancangan Sistem Informasi Akuntansi pada Pondok Pesantren Salafiyah Syafi'iyah Nurul Huda Mergosono Malang. [Thesis]. Islamic State University Maulana Malik Ibrahim, Malang. [Indonesian]

Maulana A, Bijaksana MA, Mubarok S. 2016. Perancangan Semantic Similarity based on Word Thesaurus Menggunakan Pengukuran Omiotis Untuk Pencarian Aplikasi pada I-GRACIAS. E-Proceeding of Engineering: Vol. 3(2) Agustus 2016. Telkom University. Bandung

Nuriyanti F, Izzah I, Azizah M, Yaqin M A. 2019. Business Process Model Variations in Islamic Boarding Schools Using Requirements Analysis Method. Islamic State University Maulana Malik Ibrahim, Malang. [Indonesian]

Robert AL, Davis KR. 1992. Accounting Information System. Prentice-Hall International.

THIS PAGE INTENTIONALLY LEFT BLANK