IOP Conf. Series: Materials Science and Engineering **420** (2018) 012138 doi:10.1088/1757-899X/420/1/012138

Doubleton: A Role of the Search Engine to Reveal the Existence of Relation in Information Space

M K M Nasution[1,*]

[1]Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Padang Bulan 20155 USU, Medan, Indonesia

E-mail: *mahyuddin@usu.ac.id

Abstract. By using Web, information contributes to existence of relation. The information space take a position for exploring presentation and determines the existence of relation. However, this existence depend on the access tool, the search engine. They are as interface for it. A search engine has itself behaviour. This paper express the role of search engines based on doubleton through logical reasoning and the set concept. While a similarity showed the relation is exist.

1. Introduction

Information space such as Web contain a variety of information [1]. Available information can be modeled as needed [2]. This depends not only on access tools to information spaces such as search engines, but to the treatment of information resources [3]. Treatment is needed to map the information space according to the desired result such as to reveal the existence of something [4].

In the age of industry 4.0 [5], existence is not only related to information but also relation [6, 7]. The relation of something to something else reinforces the existence of either or both, and also interdependence will trigger the presence of information about something to increase its existence [8]. Information can also be engineered to reveal the relation between something, a relation that ensures the sustainability of the existence of something [9]. Therefore, the use of search engine is as the mapping of information space so that the relations are described become one of search engine roles. This is known as doubleton.

2. Problem Definition

In principle, the relations are formed through overlapping interests [10]. The importance of something is described through its existence in the information space [11]. This principle means that if a part of the information space states something and something else, both might have a relation [12]. This is stated as co-occurrence, whereby they are to be together in one part of information [13]. The existence of such a relation is stated by each representative in the $\Omega_x \cap \Omega_y$, that is the overlap cluster of two different clusters, each of clusters represents something. Each Ω is an information space containing ω. Every ω as a webpage contains at least one or more words with which the search term t_x submitted to the search engine. This serves to raises related information. Search terms are expressed as a set of words $(w_1, w_2, \ldots, w_k)$, k is the size of t_x

or the number of words in t_x. In the query, the co-occurrence is expressed as $q = t_x, t_y$ for two different search terms [14].

In theory, the query $q = t_x, t_y$ gives the existence of a *doubleton*. A doubleton can be explained as the doubleton search term, that is $\mathcal{D} = \{\{t_x, t_y\} : t_x, t_y \in \Sigma\}$ and its vector space denoted by $\Omega_x \cap \Omega_y$ [15], a doubleton search engine event of web pages that contain a co-occurrence of t_x and t_y such that $t_x, t_y \in \omega_x$ and $t_x, t_y \in \omega_y$ [16]. The existence of doubleton is supported by singleton, which states that if there are two different search terms, $t_x \neq t_y$, $t_x, t_y \in \mathcal{S}$ where $\mathcal{S}$ is the singleton search term of search engine, there are $\Omega_x \subseteq \Omega$ and $\Omega_y \subseteq \Omega$. Ω is the space information [11]. Generally, therefore the existence of relations is expressed through the concept of similarity that has been applied to the information extraction about networks [17], but this existence needs to be proven through the relations between the word w_k, the parts of information space or webpages ω, and the search terms t. It is to determine the role that the search machine can provide in theory and implementation [18].

3. The proposed approach

To get the overlap cluster from the information space is used a query q, that is $q =$ `Mahyuddin K. M. Nasution,Opim Salim Sitompul` as an approach to match content queries with content available in the information space, where $t_x =$ `Mahyuddin K. M. Nasution` and $t_y =$ `Opim Salim Sitompul` [19]. In general this is related to the concept of probability logic as follows: A web page ω as a part of information space is relevant to a query q if the content of ω implies the query. Therefore, we use this concept to prove the existence of the relations in the information space [20].

On the other hand, the reality about it is that for every w in t, there exists w is also a member of ω, so $t \in \omega$. Therefore, the reality about it that for every w in t_x and t_y, there are w also in ω_x and ω_y such that both t_x and t_y in ω_x and ω_y [21]. This approach involves set theory to reason logically the concept of probability logic: an implication $\omega_x \wedge \omega_y \Rightarrow q$ is true.

4. Analysis and Discussion

Two clusters of the information space, i.e. Ω_x and Ω_y, are based on two implications: $\omega_x \Rightarrow t_x$ and $\omega_y \Rightarrow t_y$ are true, or

$$\Omega_x \cap \Omega_y(t_x \wedge t_y) = \begin{cases} 1 & \text{if } t_x, t_y \text{ is true at } \omega_x \wedge \omega_y,\ \omega_x, \omega_y \in \Omega \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Therefore, the inference of relation is about the probability of matching between a search term and the webpages in the information space. Some of them are match or $\rho(\omega_x \wedge \omega_y \Rightarrow q)$ is true. The logical consequence of condition $t_x \cap t_y | \omega$ whereby $(t_x, t_y) \cap \omega$ reveals that applies

$$\rho(\omega_x \wedge \omega_y \Rightarrow q) \simeq \begin{cases} 1 & \text{if } t_x, t_y \text{ is true at } \omega_x \wedge \omega_y,\ \omega_x, \omega_y \in \Omega \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

This expression as a mapping of information space around the relations depends on search term, this is the cardinality or

$$|\Omega_x \cap \Omega_y| = \sum_{\Omega} (\omega \Rightarrow t_x \wedge t_y \simeq 1) \geq \sum_{\Omega} (\omega \Rightarrow t_x \wedge t_y = 1). \tag{3}$$

whereby t_x and t_y are two search terms [14, 22, 23].

4.1. Some base concepts
Let us have two search terms t_x and t_y with an assumption that

$$t_x \cap t_y = \emptyset. \tag{4}$$

It means that $\forall w \in t_x, w \notin t_y$ and $\forall w \in t_y, w \notin t_x$. However, it is possible $\forall w \in t_x, w \in t_y$ and $\forall w \in t_y, w \notin t_x$, so that $\forall t_y \in \omega_y \exists t_y \in \omega_x$ and $\forall t_x \in \omega_x \exists t_x \in \omega_y$, or

$$\omega_x \cap \omega_y \neq \emptyset. \tag{5}$$

Thus, for two different search terms, we have

$$\Omega_x \cap \Omega_y = \{(t_x, \omega_x)_{kj}\} \cap \{(t_y, \omega_y)_{kj}\} \tag{6}$$

or

$$\begin{array}{rcl} \Omega_x \cap \Omega_y & = & \{(\{t_x, t_y\}, \{\omega_x, \omega_y\})_{kj}\} \\ & = & \{(t_x \cap t_y, \omega_x \cap \omega_y)_{kj}\} \end{array} \tag{7}$$

Based on Eq. (4) and Eq. (5), Eq. (7) be

$$\Omega_x \cap \Omega_y \neq \emptyset \tag{8}$$

or

$$|\Omega_x \cap \Omega_y| \geq 0 \tag{9}$$

Suppose $t_x \neq t_y$, it is possible that $\forall w \in t_x \exists w \in t_y$ or $\forall w \in t_y \exists w \in t_x$ so that

$$t_x \cap t_y \neq \emptyset \tag{10}$$

or it is possible that $\forall w \in t_x, w \notin t_y$ and $\forall w \in t_y \exists w \in t_x$ or vice versa, we obtain also Eq. (10). If we substitute Eq. (10) to Eq. (7), we have Eq. (8) and Eq. (9) is more than the old value. However, for $|t_y| < |t_x|$ we have that $\forall w \in t_y, w \in t_x$ and $\exists w \in t_x, w \notin t_y$ so that $\Omega_x \cap \Omega_y = \Omega_y$ or

$$|\Omega_x \cap \Omega_y| = |\Omega_y| \tag{11}$$

Thus, the value in Eq. (9) contains values based on Eq. (11). It causes a greater value of $|\Omega_x \cap \Omega_y|$.

Theoretically the results expressed by Eq. (9) are evidences of change in the existence of a relation [24].

4.2. An explanation
As an explanation, for two search terms:

(i) $t_x =$ `Mahyuddin K. M. Nasution`
(ii) $t_y =$ `Opim Salim Sitompul`

and we can use

(i) $q_1 = t_x$ for $|\Omega_x|$.
(ii) $q_2 = t_y$ for $|\Omega_y|$.
(iii) $q_3 = t_x, t_y$ for $|\Omega_x \cap \Omega_y|$.

At the time of writing, by using Google search engine we have a proof of existence about relation between two proofs of existence about something in information space, that is $|\Omega_x \cap \Omega_y| = 9,730$ whereby $|\Omega_x| = 34,100$ and $|\Omega_y| = 3,840$. While for "t_x" and "t_y", the proof of existence: the relation is $|\Omega_{"x"} \cap \Omega_{"y"}| = 274$ in information space; t_x and t_y are $|\Omega_{"x"}| = 2,240$, $|\Omega_{"y"}| = 2,860$, respectively. This shows that doubleton comparison as $|\Omega_x \cap \Omega_y| > |\Omega_{"x"} \cap \Omega_{"y"}|$ follows the singleton to prove Eq. (3) as a rule in information space, that is a role of search engine for proving the existence of relation in information space [25]. This is supported by the use of similarity measurements, for example by using Jaccard coeficient:

$$\{|\Omega_x \cap \Omega_y|, |\Omega_x|, |\Omega_y|\} \geq \{|\Omega_{"x"} \cap \Omega_{"y"}|, |\Omega_{"x"}|, |\Omega_{"y"}|\}.$$

That is

$$0.3449 > 0.0568$$

as the proof of relation existence in comparison. For example, the relation based on $\frac{78}{(378+5,160,000-78)}$ shows the relation between `SumutSiana` [26] and `TALENTA` [27].

5. Conclusion

The role of a search engine depends on mapping the information space based on query, search term, and content of webpage for proving the existence of relations between contents of webpages. The information space exemplify existence of relation based on the concept of doubleton, and it is support by their singleton.

References

[1] Waterworth J A, Lund A, and Modjeska D 2003 Experiential design of shared information spaces *Designing Information Spaces: The Social Navigation Approach*
[2] Nasution M K M 2018 Social network extraction based on Web: 1. Related superficial methods *IOP Conference Series: Materials Science and Engineering* **300(1)**.
[3] Nasution M K M and Noah S A 2017 Social Network Extraction Based on Web. A Comparison of Superficial Methods *Procedia Computer Science* **124**.
[4] Nasution M K M, Syah R, and Elveny M 2017 Studies on behaviour of information to extract the meaning behind the behaviour *Journal of Physics: Conference Series* **801(1)**
[5] Schwab K 2016 *The Fourth Industrial Revolution* Crown Business.
[6] Nasution M K M 2014 New method for extracting keyword for the social actor *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **8397 LNAI(PART 1)**
[7] Nasution M K M, Elveny M, Syah R, and Noah S A 2015 Behavior of the resources in the growth of social network *Proceedings - 5th International Conference on Electrical Engineering and Informatics: Bridging the Knowledge between Academic, Industry, and Community* (ICEEI).
[8] Nasution M K M 2016 Social network mining (SNM): A definition of relation between the resources and SNA *International Journal on Advanced Science, Engineering and Information Technology* **6(6)**
[9] Nasution M K M, Hardi M and Syah R 2017 Mining of the social network extraction *Journal of Physics: Conference Series* **801(1)**.
[10] Nasution M K M 2015 Extracting keyword for disambiguating name based on the overlap principle *Proceeding of International Conference on Information Technology and Engineering Application* (4-th ICIBA), Book 1.
[11] Nasution M K M 2012 Simple search engine model: Adaptive properties *Cornell University Library* arXiv:1212.3906 [cs.IR].
[12] Nasution M K M and Sitompul O S 2017 Enhancing extraction method for aggregating strength relation between social actors *Advances in Intelligent Systems and Computing* **573**.
[13] Elfida M, Nasution M K M, and Sitompul O S 2018 Enhancing to method for extracting social network by the relation existence *IOP Conference Series: Materials Science and Engineering* **300(1)**.
[14] Nasution M K M and Noah S A 2012 Information retrieval model: A social network extraction perspective *Proceedings - 2012 International Conference on Information Retrieval and Knowledge Management* (CAMP'12).

[15] Nasution M K M, and Noah S A 2011 Probabilistic generative model of social network based on web features *International Seminar on Operational Research* (InteriOR 2011).
[16] Nasution M K M 2012 Simple search engine model: Adaptive properties for doubleton *Cornell University Library* arXiv:1212.4702v1 [cs.IR].
[17] Nasution M K M 2011 Kolmogorov complexity: Clustering objects and similarity *Bulletin of Mathematics* **3(1)**.
[18] Nasution M K M 2018 Ontology *Journal of Physics: Conference Series*.
[19] Nasution M K M, Sitompul O S, and Noah S A 2018 Social network extraction based on Web: 3. The integrated superficial method *Journal of Physics: Conference Series* **978(1)**.
[20] Nasution M K M 2017 Modelling and simulation of search engine *Journal of Physics: Conference Series* **801(1)**.
[21] Nasution M K M, Sitompul O S, and Noah S A 2018 Social network extraction based on Web: 2. Strategies in superficial methods *Journal of Physics: Conference Series*.
[22] Nasution M K M, Syah R, and Elfida M 2018 Information Retrieval Based on the Extracted Social Network *Advances in Intelligent Systems and Computing* **662**.
[23] Elveny M, Syah R, Elfida M and Nasution M K M 2018 Information retrieval on social network: An adaptive proof *IOP Conference Series: Materials Science and Engineering* **300(1)**
[24] Nasution M K M 2018 Semantic interpretation of search engine resultant *IOP Conference Series: Materials Science and Engineering* **300(1)**.
[25] Mahyuddin K M N, Sitompul O S, Nasution S, and Ambarita H 2017 New similarity *IOP Conference Series: Materials Science and Engineering* **180(1)**.
[26] Nasution M K M 2018 SumutSiana *IOP Conference Series: Materials Science and Engineering* **309(1)**.
[27] Nasution M K M, Sitepu R, Rosmayati, Siregar M F G, Syam B , Sihombing L, Farhat, Rambe A S, Ginting B, Hasanuddin, Maulina S, Ramli, Boel T, Agustono B, Sebayang K, Amin M, Yustina I, Masfria, Zulkarnain, Setiawan, Sitompul O S, Latifah S, Sibarani R, Munir E, Tulus, Putra I B, Tala Z Z, Sari D K, Saidin, Hasibuan P M, Leviza J, Hanum H, Ginting J, Supriana T, Tarigan J, Syahrizal, Irvan, Lumbanraja P, Siregar H S, Lubis A N, Pintauli S, Nasution R O, Zulkarnain M, Purba M, Dewi H, Nasution I, Pasaribu N, Bangun P, Gea S, Thamrin H, Nasution M A, Harahap H, Nurmaini, Chahaya S I, Santosa H, Hasibuan P A Z, Khairunnisa, Marianne, Ginting E D J, Novliadi F, Eliana R, Wahyuni S E, Siregar C T, Nasution S S, Zamzami E M, Syahputra M F, Lidya M S, Hartono R, Samsuri, Iswanto A H, Suryanto D, Lubis S, Wibowo R P, Husni M, Siregar G A W, Sitorus E, Bangun H, Bukit R, Sinulingga E P, Fadli, Bakti D, Hanum C, Marsaulina I, Ambarita H, Herawati E, Onrizal, Muda I, Yuandani, Eyanoer P C, Tambun R, Rahmat R F, Siregar R, Hasugian J, Chalil D, Nasution T I, Surya I, Rosmalinda, Muchtar M A, Nasution T H, Elfida M 2018 Talenta *Journal of Physics: Conference Series*.