IOP Conf. Series: Materials Science and Engineering **420** (2018) 012137 doi:10.1088/1757-899X/420/1/012137

Singleton: A Role of the Search Engine to Reveal the Existence of Something 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. Information is very important to the existence of something. The information space in this case becomes a continuously explored presentation and determines the existence of something. However, that existence must be related to the access tool, the search engine, serving as an intermediary. Not a few search engines have been present, each having different behaviour. This paper reveals the role of search engines based on singleton through logical reasoning and the set concept. While some measurement results have showed how the concept is played by search engines.

1. Introduction

In such a vast space of information with an indeterminate dimension, like the Web, a tool is needed to map information [1]. Mapping information related to something in the industry era 4.0 is to prove the existence of something [2]. Furthermore, information based on data so far has grown no longer following or exceeding the exponential law [3]. As a result, research on something is not sufficiently dealing with the sample only, but the large data that show the behavior of something generally and this behavior will represent the characteristics of the population [4].

The search engine is the most commonly used access tool for retrieving the related information based on its mapping capability [5]. Therefore, every search engine has its weaknesses and advantages [6], i.e. they are vary. Faced with large data, each search engine has the special behavior, although there is their general behavior [7]. However, from the view point of designing and developing the search engine, the initial characteristics need to be expressed from the existing search engines. This paper intends to reveal some search engine properties semantically in occurrence so as to state the role of search engines in expressing something in what we call as the singleton.

2. Problem Definition

Any search engine in general will present briefly and solidly information about something based on its mapping from the information space Ω [8]. For example, for query $q =$ `Mahyuddin K. M. Nasution` the search engine: Google[1] generates 30,400 hits, Yahoo[2] generates 6,750 hits,

[1] https://www.google.co.id

[2] https://id.search.yahoo.com

Geevv[3] generates 147,000 hits, etc [9, 10]. Search engines generally provide information about the number of query contents contained in the information space as well as with documents related to the content of the query [11]. However, the contents in query are sorted from the most appropriate [7].

Suppose we express a search term as follows [12].

T1 A *search term* is a *term* consists of one or more words. We notify the word as w and the search term as $t = (w_1 w_2 \cdots w_k)$ whereby $|t| = k$ is size of t, l is a number of vocabularies (tokens) in t_x, and $k \geq l$.

Suppose ω is the document consists relatively of a set of words or $\omega = \{w_i | i = 1, \ldots, I\}$. Any document becomes an element in the information space or $\omega \in \Omega$, $j = 1 \ldots, J$. Thus it can be stated that a search engine as follows [13].

T2 $\Omega = \{(t, \omega)_{k,j}\}$ is a pair describes the relationship between a term and a document by the search engines. $\{\omega_j | j = 1, \ldots, J\}$ is a set of documents index by search engine for Ω.

In information space Ω, let us state ω for webpages, then the information space is Web. If the query contains a search term or $q = t$, the relationship in T2 states webpage ω is relevant to a query q that can be expressed as a form of inference, i.e. $\omega \Rightarrow q$, meaning this implication is true [14]. So,

$$\omega \Rightarrow q = \begin{cases} 1 & \text{if } t \text{ is true at } \omega \in \Omega \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In this case, the search term exactly as they appear on webpages or webpages contain search terms (expressed in "t") exactly the same as the query [15, 16, 17]. If the webpage is more than one, the list of webpages presented is depend on the number of search term in a webpage, $n(t)$ $t \in \omega$. However, inference can also be constructed based on the probability of matching the content of the document and the search term, i.e. the implication $\omega \Rightarrow q$ is true if some of it matches or $\rho(\omega \Rightarrow q)$ is true. The logical consequence of condition $t|\omega$ by which $t \cap \omega$ reveals that applies

$$\rho(\omega \Rightarrow q) \simeq \begin{cases} 1 & \text{if } t \text{ is true at } \omega \in \Omega \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

In general this is expressed as cardinality of mapping of information space depends on search term or

$$|\Omega_x| = \sum_{\Omega} (\omega \Rightarrow t_x \simeq 1) \geq \sum_{\Omega} (\omega \Rightarrow t_x = 1), \tag{3}$$

whereby t_x is any search term and $\Omega(\omega \Rightarrow t_x)$ formalizes the reality of the statement that if t is true at $\omega \in \Omega$. Thus, if there are two search terms t_x and t_y under different conditions, there is the different relationships between $|\Omega_x|$ and $|\Omega_y|$.

3. The proposed approach

First of this approach we define the singleton as follows [13]

D1 A singleton Ω_x is a search engine event of web pages that contain an occurrence of $t_x \in \omega_x$, $|\Omega_z|$ is as the cardinality of Ω_x.

Second, we will prove theoretically that for every t_x and t_y as search terms holds characteristics as follows.

L1 *If* $t_y \neq t_y$ *and* $t_y \cap t_x = \emptyset$, *then* $\Omega_y \cap \Omega_x = \emptyset$.

[3] https://geevv.com

L2 *If* $t_x \neq t_y$, $t_x \cap t_y \neq \emptyset$ *and* $|t_y| < |t_x|$, *then singleton of* t_x *and* t_y *has a relation as* $\Omega_x = \Omega_x \cup \Omega_y$.

L3 *If* $t_x \neq t_y$, $t_x \cap t_y = \emptyset$, *and* $\omega_x \cap \omega_y \neq \emptyset$, *then* $\Omega_x \cap \Omega_y \neq \emptyset$.

Further downgrading some other properties in theory and explaining the existence of something as well as the existence of SumutSiana [18] or TALENTA [19].

4. Analysis and Discussion

By involving logical reasoning, the analysis is done in theory based on the concept of the set [13], then gives the examples as the explanation [12].

4.1. Some derived concepts

$\forall t_x$ and $\forall t_y$, based on T1 $t_x \neq t_y$ and $t_x \cap t_y = \emptyset$ mean that for all $w_x \in t_x$ and for all $w_y \in t_y$ we have $w_x \notin t_y$ and $w_y \notin t_x$ so that for all $w_x \in \omega_x$ and for all $w_y \in \omega_y$ we obtain $w_x \notin \omega_y$ and $w_y \notin \omega_x$. Thus

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

and

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

Based on T2, two independent events $\Omega_x = \{(t_x, \omega_x)_{kj}\}$ and $\Omega_y = \{(t_y, \omega_y)_{kj}\}$ therefore be

$$\Omega_x \cap \Omega_y = \emptyset. \tag{6}$$

P1 *If* $\Omega_x \neq \Omega_y$, *then* $|\Omega_x \cup \Omega_y| = |\Omega_x| + |\Omega_y|$.

In other words, for search terms t_x and t_y in L1 we obtain

$$t_x \cup t_y = t_y \cup t_x \tag{7}$$

and

$$\omega_x \cup \omega_y = \omega_y \cup \omega_x \tag{8}$$

and by combining all meaning of Eqs. (4), (5), (6), (7), and (8) we derive $\{(t_x, \omega_x)_{kj}\} \cup \{(t_y, \omega_y)_{kj}\} = \Omega_x \cup \Omega_y$ and

$$|\Omega_x \cup \Omega_x| = |\Omega_x| + |\Omega_y| \tag{9}$$

Assumption in L2 means that for all w_y in t_y, w_y in t_x, but exist w_x in t_x, w_x is not in t_y, then for all w_y in ω_y, w_y in ω_x, but exist w_x in ω_x, w_x is not in ω_y such that

$$t_x \cap t_y = t_y \tag{10}$$

and

$$t_x \cup t_y = t_x, \tag{11}$$

or

$$\omega_x \cap \omega_y = \omega_y \tag{12}$$

and

$$\omega_x \cup \omega_y = \omega_x. \tag{13}$$

Therefore, $\Omega_x = \{(t_x, \omega_x)_{kj}\}$, we apply Eq. (11) and Eq. (13) and we obtain $\Omega_x = \{(t_x, \omega_x)_{kj}\} = \{(t_x \cup t_y, \omega_x \cup \omega_y)_{kj}\} = \{(t_x, \omega_x)_{kj} \cup (t_y, \omega_y)_{kj}\} = \{(t_x, \omega_x)_{kj}\} \cup \{(t_y, \omega_y)_{kj}\} = \Omega_x \cup \Omega_y$, or

$$\Omega_x = \Omega_x \cup \Omega_y. \tag{14}$$

P2 *If* Eq. (12) *and* Eq. (13) *be in effect based on* Eq. (10) *and* Eq. (11), *then* $|\Omega_x| = |\Omega_x|+|\Omega_y|$.

Based on implication of Eq. (6) to (14) we obtain $|\Omega_x| = |\Omega_x \cup \Omega_y| = |\Omega_x| + |\Omega_y| - |\Omega_x \cap \Omega_y| = |\Omega_x| + |\Omega_y| - |\emptyset|$ or

$$|\Omega_x| = |\Omega_x| + |\Omega_y|. \tag{15}$$

Assumption L3 means that for all search term t_x and t_y if for every w_x in t_x, w_x is not in t_y, and for all w_y in t_y, w_y is not in t_x, then Eq. (4) and Eq. (7) be in effect, but for all w_x in ω_x the w_x also is in ω_y and for all w_y in ω_y the w_y also is in ω_x, or

$$\omega_x \cap \omega_y = \omega_x = \omega_y \tag{16}$$

and

$$\omega_x \cup \omega_y = \omega_x = \omega_y \tag{17}$$

Therefore, we conclude that

$$\Omega_x \cap \Omega_y \neq \emptyset. \tag{18}$$

As description, for this case let $w_1 = aaa \in t_y$ and $w_2 = aaab \in t_x$, and we can define that w_2 approach to w_1 or $w_1 \simeq w_2$ so that $t_x \simeq t_y$.

L4 *If* $w_1 \simeq w_2$, *then* $t_1 \simeq t_2$.

P3 *If* L4 *be in effect, then* $|\Omega_x| = |\Omega_y|$.

For $\Omega_x = \{(t_x, \omega_x)_{kj}\}$ and $\Omega_y = \{(t_y, \Omega_y)_{kj})\}$ we apply Eq. (16) and we obtain

$$\begin{array}{rcl}
\{(t_x,\omega_x)_{kj}\} \cap \{(t_y,\omega_y)_{kj}\} &=& \{(t_x,\omega_x)_{kj}\} \cap \{(t_y,\omega_y)_{kj}\} \\
\{(t_x,\omega_x)_{kj}\} \cap \{(t_y,\omega_y)_{kj}\} &=& \Omega_x \cap \Omega_y \\
\{(t_x,\omega_x)_{kj}\} \cap \{(t_y,\omega_x)_{kj}\} &=& \Omega_x \cap \Omega_y \\
\{(t_x,\omega_x)_{kj}\} \cap \{(t_y \simeq t_x,\omega_x)_{kj}\} &=& \Omega_x \cap \Omega_y \\
\{(t_x,\omega_x)_{kj}\} \cap \{(t_x,\omega_x)_{kj}\} &=& \Omega_x \cap \Omega_y \\
\{(t_x,\omega_x)_{kj}\} &=& \Omega_x \cap \Omega_y \\
\Omega_x &=& \Omega_x \cap \Omega_y \\
.
\end{array} \tag{19}$$

Similarly, we have also

$$\Omega_y = \Omega_x \cap \Omega_y. \tag{20}$$

In other words,

$$\begin{array}{rcl}
|\Omega_x| &=& |\Omega_x \cap \Omega_y| \\
.|\Omega_x| &=& |\Omega_y| \\
.
\end{array} \tag{21}$$

4.2. Towards explanation

Web pages within the information space have been delineated by various search engines. The role of the search engine has been demonstrated by the logic language and the set language for the concept of word w, the concept of search term t, and the concept of the document or web page ω. The combination of each relations: L1, L2, L3 and L4 with their derivatives P1, P2, and P3 provided different characteristics that change the role of the search engine whereby their measurements can by using what is called the singleton. Thus, the role of the search engine can be altered follows the interests by understanding the function of that characteristic.

As an explanation the review is done on the case of a set of search terms in {`Mahyuddin K. M. Nasution`, `Mahyuddin Nasution`, `Mahyuddin`} by involving the search engines: {Google,

Table 1. Singleton of search engine resultant

Search term (t & "t")	Resultant of Search Engine in $|\Omega|$ Google	Yahoo	Geevv
`Mahyuddin K. M. Nasution`	28,200	6,700	145,000
`Mahyuddin Nasution`	62,000	15,100	8,660
`Mahyuddin`	538,000	168,000	39,400
"`Mahyuddin K. M. Nasution`"	1.830	619	1,850
"`Mahyuddin Nasution`"	1,050	221	57
"`Mahyuddin`"	592,000	95,300	39,400

Yahoo, Geevee}, with which their resultants respectively are shown in Table 1. In general all search engines have a way of mapping different information spaces. For t_x = `Mahyuddin K. M. Nasution`, t_y = `Mahyuddin Nasution`, and t_z = `Mahyuddin`, though $|t_x| \geq |t_y| \geq |t_z|$, the resultant generated by the first two search engines shows the uniformity that each singleton is in reverse in sequence, but this is not followed by a singleton with query content containing the appropriate pattern to queries differently. Thus the existence of something is indicated by a search engine based on the size of the search term. The existence of something in the information space is more guaranteed to exist than naming involves one word. However, because of the terms used like SumutSiana $|\Omega_{SumutSiana}| = 383$ (in one word and at the time of this writing), this shows the existence of the SumutSiana and the concept as something new, while the concept of TALENTA $|\Omega_{TALENTA}| = 5,180,000$ (at the time of this writing) require redefinition in accordance with its function, i.e. **T**ROPICAL SCIENCE AND MEDICINE, **A**GROINDUSTRY, **L**OCAL WISDOM, **E**NERGY (SUSTAINABLE), **N**ATURAL RESOURCES (BIODIVERSITY, FOREST, MARINE, MINE, TOURISM), **T**ECHNOLOGY (APPROPRIATE) and **A**RTS (ETHNIC) [18].

5. Conclusion

The role that the search engine plays in mapping the information space is directly related to the context of the search term. The impact of search terms on search engine roles is based on logical reasons, but this is revealed through the treatment of search terms content and web pages, that is to the order and number of search terms in the document.

References

[1] Abrams D, Baecker R, and Chignell M 1998 Information archiving with bookmarks: Personal web space construction and organization *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*
[2] Schwab K 2016 *The Fourth Industrial Revolution* Crown Business.
[3] 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)**
[4] 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).
[5] 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**.
[6] Nasution M K M, Hardi M and Syah R 2017 Mining of the social network extraction *Journal of Physics: Conference Series* **801(1)**.
[7] Nasution M K M 2017 Modelling and simulation of search engine *Journal of Physics: Conference Series* **801(1)**.
[8] Nasution M K M 2018 Semantic interpretation of search engine resultant *IOP Conference Series: Materials Science and Engineering* **300(1)**.
[9] Matsuo Y, Tomobe H, and Nishimura T 2007 Robust estimation of Google counts for social network extraction *Proceedings of the National Conference on Artificial Intelligence.*

[10] Adamic L A, Zhang J, Bakshy E, and Ackerman M S 2008 Knowledge sharing and yahoo answers: Everyone knows something *Proceedings of the 17th international conference on World Wide Web.*

[11] 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)**

[12] 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)**

[13] Nasution M K M 2012 Simple search engine model: Adaptive properties *Cornell University Library* arXiv:1212.3906 [cs.IR].

[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 2018 Social network extraction based on Web: 1. Related superficial methods *IOP Conference Series: Materials Science and Engineering* **300(1)**.

[16] Nasution M K M and Noah S A 2017 Social network extraction based on Web. A comparison of superficial methods *Procedia Computer Science* **124**.

[17] 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.*

[18] Nasution M K M 2018 SumutSiana *IOP Conference Series: Materials Science and Engineering* **309(1)**.

[19] 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.*