IOP Conf. Series: Materials Science and Engineering **300** (2018) 012053 doi:10.1088/1757-899X/300/1/012053

Semantic interpretation of search engine resultant

M K M Nasution

Teknologi Informasi, Fasilkom-TI, Universitas Sumatera Utara, Padang Bulan 20155 USU Medan Indonesia

E-mail: mahyuddin@usu.ac.id

Abstract. In semantic, logical language can be interpreted in various forms, but the certainty of meaning is included in the uncertainty, which directly always influences the role of technology. One results of this uncertainty applies to search engines as user interfaces with information spaces such as the Web. Therefore, the behaviour of search engine results should be interpreted with certainty through semantic formulation as interpretation. Behaviour formulation shows there are various interpretations that can be done semantically either temporary, inclusion, or repeat.

1. Introduction

In the logic of language there are issues that present the motivation for fomulating the phenomena of relationships between entities [1, 2]. The phenomena of this relationship is recognized as semantics such as meronymy [3], holonomy [8], hyponymy [5], synonymy [6], and polysemy [7]. Although the characteristics of this different relationships initially relate only to natural language, but because the language logically runs the technology, the consequences of any technology are directly influenced by language behavior [8, 9, 10]. For example, the disambiguation of entity name not only becomes the social nature of the language [11], but also the special nature of search engine for the social media, like in Web [12].

In general, researchers involve the natural language processing (NLP) and the search engine (SE) to reveal the semantic behavior of language logic [13]. This is done by affirming language needs such as keywords, alias search, grammar, and data structure [14]. But not a lot of efforts to explore the potential of semantic formulation based on the results were returned by a search engine, the resultant. Each search engine gives characteristics to the information it produces [15]. Therefore, search engine resultants become important to interpret through implementation of set theory of the words. This paper will describe some semantic behavior formulations generated by search engines from the information space it accesses.

2. Related Works and Motivation

Search engine resultant is the result of return of each search engine based on query content [16, 17]. Query content is a search term, and the concept of return results is generally based on the similarity between search terms and the content of the information space with which the search engine becomes its interface [18]. If $t = (w_1, w_2, \dots, w_n)$ is a search term with w_i, $i - 1, \dots, I$ are the words. Suppose Ω is information space with members are in $\{\omega_1, \omega_2, \dots, \omega_J\}$ with which every ω_j consists of the words w_i, $k = 1, \dots, I$, and size of t are smaller or equal to

IOP Conf. Series: Materials Science and Engineering **300** (2018) 012053 doi:10.1088/1757-899X/300/1/012053

Table 1. Similarity of query and contents of pages in information space

t_i	$\|t\|$	$\|\omega\|$	kecocokan dengan ω	$\|t \in \omega\|$	sim
t_1	5	5	5 same words	5	1.00
t_2	5	5	4 same words	4	0.64
t_3	5	5	3 same words	3	0.36
t_4	5	5	2 same words	2	0.16
t_5	5	5	1 same words	1	0.04
t_6	5	5	0 same words	0	0.00

the size of ω or $|t| \leq |\omega|$ [19, 20]. Suppose that the search term size n with the member of ω gives the largest pseudo size is n, thus based on the measurement of similarity [21]

$$sim = \frac{2|xy|}{|x|^2 + |y|^2} \tag{1}$$

we get Table 1 for $|t| = 5$.

Suppose to apply the similarity, the search engine is modeled as follows [22, 23, 24].

Definition 1. Let Ω contains the ordered pair of the terms t_i and the pages ω_j, or (t_i, ω_j), $i = 1, \ldots, I$, $j = 1, \ldots, J$. (t_i, ω_j) as a relation table of t_i and ω_j such that $\Omega = \{(t, \omega)_{ij}\}$ as presentation of search engine, and $|\Omega|$ is a cardinality of Ω.

Definition 2. Let t_x is a search term and $t_x \in \mathcal{S}$ is a set of singleton search term of search engine. A vector space $\Omega_x \subseteq \Omega$ is a singleton search engine event (or singleton space of event) of pages that contain an occurrence of $t_x \in \omega_x$. The cardinality of Ω_x is denoted by $|\Omega_x|$, $|\Omega_x| \leq |\Omega|$.

In occurrence, the semantic interpretation of Table 1 describes that for a query $q = t_x$, the search engine returns results sequentially based on the correspondence between t_x and contents of information space. In other words, first the page in information space is relevant to a query in logical implication if it implies the query or if $\omega \Rightarrow q$ is true: $\omega \Rightarrow t$ is true for all $\omega \in \Omega$, simply $\omega \Rightarrow t = 1$ [25]. In other hand, for $|t_x| = k$ vocabularies we have $2^k - 2$ anothers of $\{\{t_k^{2^k-2}\} \subset \{w_1, w_2, \ldots, w_k\} = t_k\}$. Thus, $\omega \Rightarrow q$ is true based on Eq. (1) or simply

$$\omega \overset{sim}{\Longrightarrow} q = \omega \overset{sim}{\Longrightarrow} t_k \tag{2}$$

whereby if each the matched content of pages we denoted as search terms also: $t_x, t_y, \ldots, t_z$ then $t_z \subset \ldots \subset t_y \subset t_x$. Therefore, each search engine has a meaning semantics as follows [22].

Theorem 1. Let $t_x, t_y \in \mathcal{S}$ are two different search terms. If Ω_x and Ω_y are the singleton search engine events of t_x and t_y, respectively, then

$$|\Omega_x| = |\Omega_x| + |\Omega_y| \tag{3}$$

Definition 3. Let $t_x, t_y \in \mathcal{S}$, $t_x \neq t_y$. A doubleton search term is $\mathcal{D} = \{\{t_x, t_y\} : t_x, t_y \in \mathcal{S}\}$ and its vector space denoted by $\Omega_x \cap \Omega_y$ is a double search engine event (or doubleton space of eent) of pages that contain a co-occurrence of t_x and t_y such that $t_x, t_y \in \omega_x$ and $t_x, t_y \in \omega_y$ where $\Omega_x, \Omega_y, \Omega_x \cap \Omega_y \subseteq \Omega$.

IOP Conf. Series: Materials Science and Engineering **300** (2018) 012053 doi:10.1088/1757-899X/300/1/012053

Similar to occurrence, in co-occurrence each search engine based on doubleton has a meaning semantics as follows [23].

Theorem 2. Let $t_x, t_y \in \mathcal{S}$, $t_x \neq t_y$. If $t_x, t_y \in \Omega_x \cap \Omega_y$, then

$$|\Omega_x \cap \Omega_y| = |\Omega_x \cap \Omega_y| + |\Omega_x \cap \Omega_x| + |\Omega_y \cap \Omega_y| \tag{4}$$

3. An Approach

As adaptive approach we propose some scenarios to phenomena of implementation about set of words. Based on assumption that $t_x, t_y \in \mathcal{S}$ we get three concepts [22]

(i) $t_x \neq t_y$ and $q_1 = t_x$ and $q_2 = t_y$, or two search terms in different queries.
(ii) $t_x \neq t_y$, $t_x \cap t_y = \emptyset$, and $\omega_x \cap \omega_y \neq \emptyset$.
(iii) $t_x \neq t_y$, $t_x \cap t_y \neq \emptyset$ and $|t_y| < |t_x|$.

Lemma 1. If $q_x = t_x$ and $q_y = t_y$, then

$$|\Omega_x \cap \Omega_y| = 0 \tag{5}$$

Proof. For queries q_x and q_y we have $\Omega_x = \{(t_x, \omega_x)_{ij}\}$ and $\Omega_y = \{(t_y, \omega_y)_{ij}\}$, respectively. However, $\Omega_x \cap \Omega_y = \{(t_x, \omega_x)_{ij}\} \cap \{(t_y, \omega_y)_{ij}\} = \{(t_x \cap t_y, \omega_x \cap \omega_y)\}$ or based on Eq. (2) $\Omega_x \cap \Omega_y = \{(q_x \cap q_y, \omega_x \cap \omega_y)\} = \{\omega_x \cap \omega_y \Rightarrow q_x \cap q_y)\} = \{(\omega_x \Rightarrow q_x) \wedge \omega_y \Rightarrow q_y)\} = \emptyset$. In other words, although possible that $t_x \cap t_y \neq \emptyset$ and then not $\Omega_x \cap \Omega_y \neq \emptyset$ because resultants of queries q_1 and q_2 are different, clearly that $|\Omega_x \cap \Omega_y| = 0$.

Lemma 2. Let $t_x, t_y \in \mathcal{S}$ are two terms. If $t_x \neq t_y$, $t_x \cap t_y = \emptyset$, and $\omega_x \cap \omega_y \neq \emptyset$, then $\Omega_x = \Omega_y$, $\Omega_x, \Omega_y \subseteq \Omega$.
Proof. Based on assumption, by Definition 1 and Definition 2 we have $\forall w_x \in t_x$, $w_x \notin t_y$, and $\forall w_y \in t_y$, $w_y \notin t_x$, then $t_x \cap t_y = \emptyset$ or $t_x \cup t_y = t_y \cup t_x$, but $\forall w_x \in \omega_x$, $w_x \in \omega_y$ and $\forall w_y \in \omega_y$, $w_y \in \omega_x$, then $\omega_x \cap \omega_y = \omega_x = \omega_y$, $\omega_x \cup \omega_z = \omega_z \cup \omega_x = \omega_x = \omega_z$. For $\Omega_x = \{(t_x, \omega_x)\} = \{(t_x, \omega_x \cap \omega_y)\} = \{(t_x, \omega_x)\} \cap \{(t_x, \omega_y)\} = \{(t_x, \omega_x)\} \cap \{(t_y, \omega_y)\} = \Omega_x \cap \Omega_y$ we get $\Omega_x = \Omega_x \cap \Omega_y$. Similarly, we get $\Omega_y = \Omega_y \cap \Omega_x$. Thus, $\Omega_x = \Omega_y$ is proved.

Proposition 1. Let $t_x, t_y \in \mathcal{S}$ are two terms. If $t_x \neq t_y$ and $t_x \cap t_y \neq \emptyset$, then $|\Omega_x \cup \Omega_y| = |\Omega_x| + |\Omega_y|$.
Proof. Based on assumption, by Definition 1 and Definition 2 we have $\forall w_x \in t_x, \exists w_x \in t_y$ and $\forall w_y \in t_y$, $\exists w_y \in t_x$, then $\forall w_x \in \omega_x, \exists w_x \in \omega_y$ and $\forall w_y \in \omega_y$, $\exists w_y \in \omega_x$ such that $\{(t_x, \omega_x)\} \cap \{(t_y, \omega_y)\} = \Omega_x \cap \Omega_y \subset \Omega_x = \{(t_x, \omega_x)\}$ or $\Omega_x \cap \Omega_y \subset \Omega_y$ and $\{(t_x, \omega_x)\} \cup \{(t_y, \omega_y)\} = \Omega_x \cup \Omega_y \supset \Omega_x$ or $\Omega_x \cup \Omega_y \supset \Omega_y$. Thus, $|\Omega_x \cup \Omega_y| = |\Omega_x| + |\Omega_y| - |\Omega_x \cap \Omega_y|$, and based on Lemma 1 and Lemma 2, we have $|\Omega_x \cup \Omega_y| = |\Omega_x| + |\Omega_y|$.

Lemma 3. Let $t_x, t_y \in \mathcal{S}$ are terms. If $t_x \neq t_y$, $t_x \cap t_y \neq \emptyset$ and $|t_y| < |t_x|$, then

$$\Omega_x = \Omega_x \cup \Omega_y \tag{6}$$

Proof. Based on assumption, by Definition 1 and Definition 2, we have $\forall w_y \in t_y, w_y \in t_x$, $\exists w_x \in t_x, w_x \notin t_y$ and then $\forall w_y \in \omega_y$, $w_y \in \omega_x$, $\exists w_x \in \omega_x$, $w_x \notin \omega_y$ such that $t_x \cap t_y = t_y$ and $t_x \cup t_x = t_x$ or $\omega_x \cap \omega_y = \omega_y$ and $\omega_x \cup \omega_y = \omega_x$. Let $\Omega_x = \{(t_x, \omega_x)\}$, we have $\Omega_x = \{(t_x, \omega_x)\} = \{(t_x \cup t_y, \omega_x \cup \omega_y)\} = \{(t_x, \omega_x) \cup (t_y, \omega_y)\} = \{(t_x, \omega_x)\} \cup \{(t_y, \omega_y)\} = \Omega_x \cup \Omega_y$. It is proved.

IOP Conf. Series: Materials Science and Engineering **300** (2018) 012053 doi:10.1088/1757-899X/300/1/012053

The next step we use the concept of doubleton [23] as follows.

Proposition 2. Let $t_x, t_y \in \mathcal{S}$ are search terms. If $t_x \neq t_y$, but $\{(t_x, \omega_x)\} \cap \{(t_y, \omega_y)\} \neq \emptyset$, then for a doubleton search engine event of t_x and t_y is the $\Omega_x \cap \Omega_y$, and $\Omega_x, \Omega_y \subseteq \Omega$,

$$|\Omega_x \cap \Omega_y| \leq |\Omega_x| \leq |\Omega| \tag{7}$$

and

$$|\Omega_x \cap \Omega_y| \leq |\Omega_y| \leq |\Omega|. \tag{8}$$

Proof. By using assumption of Lemma 3, $\Omega_x \cap \Omega_y = \{(t_x, \omega_x)\} \cap \{(t_y, \omega_y)\} = \{(t_x \cap t_y, \omega_x \cap \omega_y)\} = \{(t_y, \omega_y)\} = \Omega_y$, or simply

$$\Omega_x \cap \Omega_y = \Omega_y, \tag{9}$$

or for $\Omega_y \subset \Omega_x$, we obtain $\Omega_x \cap \Omega_y \subset \Omega_x$. While based on assumption of Proposition 1, $\Omega_x \cap \Omega_y = \{(t_x, \omega_x)\} \cap \{(t_y, \omega_y)\} = \{(t_x \cap t_y, \omega_x \cap \omega_y)\} = \{(\emptyset, \emptyset)\} = \emptyset$. It means that

$$\Omega_x \cap \Omega_y \subset \Omega_x \tag{10}$$

and

$$\Omega_x \cap \Omega_y \subset \Omega_y \tag{11}$$

By using assumption of Lemma 2, $\Omega_x \cap \Omega_y = \{(t_x, \omega_x)\} \cap \{(t_y, \omega_y)\} = \{(t_x \cap t_y, \omega_x \cap \omega_y)\} = \{(t_x, \omega_x)\} = \Omega_x$, or simply

$$\Omega_x \cap \Omega_y = \Omega_x \tag{12}$$

Thus, Eqs. (9), (10), (11), and (12) give $|\Omega_x \cap \Omega_y| \leq |\Omega_x| \leq |\Omega|$ or $|\Omega_x \cap \Omega_y| \leq |\Omega_y| \leq |\Omega|$.

4. Proof of Theorem

Let Ω_x and Ω_y are two singletons, by using Eq. (5) in Lemma 1 [22], Eq. (6) of Lemma 3 be

$$\begin{aligned} |\Omega_x| &= |\Omega_x \cup \Omega_y| \\ &= |\Omega_x| + |\Omega_y| - |\Omega_x \cap \Omega_y| \\ &= |\Omega_x| + |\Omega_y| \end{aligned} \tag{13}$$

and Theorem 1 is proved.

From Eq. (12) [23] we get

$$\begin{aligned} |\Omega_x \cap \Omega_y| &= |\Omega_x| & \\ &= |\Omega_x| + |\Omega_y & \text{Lemma 1} \\ &= |\Omega_x| + |\Omega_x \cap \Omega_y| & \text{Eq. (9)} \\ &= |\Omega_x| + |\Omega_y| + |\Omega_x \cap \Omega_y| & \text{Lemma 1} \\ &= |\Omega_x \cap \Omega_y| + |\Omega_x \cap \Omega_x| + |\Omega_y \cap \Omega_y| & \end{aligned} \tag{14}$$

and Theorem 2 is proved.

Proposition 3. Let $t_z, \ldots, t_y, t_x \in \mathcal{S}$ are terms where $t_z \neq \ldots \neq t_y \neq t_x$ and $|t_z| < \ldots |t_y| < |t_x|$, then $\Omega_x = \Omega_x \cup \Omega_y$ holds recursively.

Proof. By the Lemma 3 and an assumption that $|t_x| < \ldots |t_y| < |t_x|$, then for $|t_z| < |t_{z_1}|$ we have $\Omega_{z_1} = \Omega_{z_1} \cup \Omega_z$, for $|t_y| < |t_x|$ we have $\Omega_x = \Omega_x \cup \Omega_y$. Thus, we obtain $\Omega_x = \Omega_x \cup \Omega_y \cup \ldots \cup \Omega_z$, or $\Omega_x = \Omega_x \cup \Omega_y$ be recursive where $\Omega_y \cup \ldots \cup \Omega_z$ is a part of Ω_x.

Corollary 1. Let $t_z, \ldots, t_y, t_x \in \mathcal{S}$. If $t_z \neq \ldots \neq t_y \neq t_x$ and $|t_z| < \ldots < |t_y| < |t_x|$, then

$$|\Omega_x| = |\Omega_x| + |\Omega_y| + \ldots + |\Omega_z| \tag{15}$$

Proof. Based on Theorem 1 and Proposition 2. we have $|\Omega_x| = |\Omega_x| + |\Omega_y| = |\Omega_x| + |\Omega_y \cup \ldots| = |\Omega_x| + |\Omega_y| + \ldots = |\Omega_x| + |\Omega_y| + |\ldots \cup \Omega_z| = |\Omega_x| + |\Omega_y| + \ldots + |\Omega_z|$.

Corollary 2. Let $t_x, t_y, t_a, t_b \in \mathcal{S}$, $t_x \cap t_y = \emptyset$ and $t_a \cap t_b \neq \emptyset$. If $|\Omega_x| = |\Omega_x| + |\Omega_a|$ and $|\Omega_y| = |\Omega_y| + |\Omega_b|$, then $|\Omega_x \cap \Omega_y| \geq 0$.
Proof. This is a direct consequence of Proposition 3 and Lemma 3.

Corollary 3. Let $t_z, \ldots, t_y, t_x \in \mathcal{S}$. If $t_z \neq \ldots \neq t_y \neq t_x$ and $|t_z| < \ldots < |t_y| < |t_x|$, then Eq. (14) holds recursively.
Proof. This is a direct consequence of Corollary 1, i.e

$$|\Omega_x \cap \Omega_y| = |\Omega_x \cap \Omega_y| + |\Omega_x \cap \Omega_x| + |\Omega_y \cap \Omega_y| + |\Omega_y \cap \ldots| + \ldots + |\ldots \cap \Omega_z| + |\Omega_z \cap \Omega_z| \tag{16}$$

5. Conclusion

Interpreting the resultant of search engines is semantically into some formulations is to provide certainty of the relationship between several entities, and it produces a temporary behavior of certainty for the events with changes of information dynamically, the inclusion of other information within an available information, and the repeat information recursively as much as iterations possible. By doing so, it is systematically possible to screen uncertain traits for trusty information relate on other and future works.

References

[1] A A Ferreira, M A Goncalves, and A H F Laender 2012 A brief survey of automatic methods for author name disambiguation *ACM SIGMOD Record* **41(2)**.
[2] M K M Nasution, and O S Sitompul 2017 Enhancing extraction method for aggregating strength relation between social actors *Advances in Intelligent Systems and Computing* **573**.
[3] H T Nguyen and T H Cao 2008 Named entity disambiguation on an ontology enriched by Wikipedia *IEEE*.
[4] M K M Nasution 2015 Extracting keyword for disambiguating name based on the overlap principle *International Conference on Information Technology and Engineering Application* (4-th ICIBA), Book 1. (arXiv: 1602.00104v1 [cs:IR] 30 Jan 2016).
[5] H T Nguyen and T H Cao 2008 Named entity disambiguation: A hybrid statistical and rule-based incremental approach *ASWC*, LNCS **5367**.
[6] L Lloyd, V Bhagwan, D Gruhl, and A Tomskins 2005 Disambiguation of references to individuals *RJI0364 Computer Science*.
[7] Y Song, J Huang, I G Councill, J Li, and C L Giles 2007 Generative models for name disambiguation (poster paper) *WWW 2007*.
[8] M K M Nasution, M Elveny, R Syah, and S A Noah 2015 Behaviour of the resoures in the growth of social network *Proceedings - 5th International Conference on Electrical Engineering and Informatics: Bridging the Knowledge between Academic, Industry, and Community, ICEEI 2015 7352551*.
[9] M K M Nasution, M Hardi, R Sitepu 2016 Using social networks to assess forensic of negative issues *Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016 7577513*.
[10] M K M Nasution, M Hardi, R Sitepu, and E Sinulingga 2017 A method to extract the forensic about negative issues from Web *IOP Conference Series: Materials Science and Engineering* **180(1)**, 012241.
[11] H Schutze 1998 Automatic word sense discrimination *Computational Linguistics* **24(1)**.
[12] R Bekkerman and A McCallum 2005 Disambiguating Web appearances of people in a social network *The Fourteenth International World Wide Web Conference or Proc. of WWW*.
[13] M K M Nasution, R Syah and M Elfida 2018 Information retrieval based on the extracted social network *Applied Computational Intelligence and Mathematical Methods*, Advances in Intelligent Systems and Computing **662**.

[14] M K M Nasution 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).
[15] M K M Nasution 2017 Modelling and simulation of search engine *Journal of Physics: Conference Series* **801(1)**, 012078.
[16] M K M Nasution 2013 Superficial method for extracting academic sosial network from the Web *Ph.D Dissertation, Universiti Kebangsaan Malaysia*.
[17] Aaron, O S Sitompul, and R F Rahmat 2015 Distributed autonomous Neuro-Gen Learning Engine for content-based document file type identification, *IEEE International Conference on Cyber and IT Service Management*.
[18] M K M Nasution, 2016, Social network mining: A definition of relation between the resources and SNA, *International Journal on Advanced Science, Engineering and Information Technology* **6(6)**.
[19] M K M Nasution, and S A Noah 2011 Extraction of academic social network from online database *2011 International Conference on Semantic Technology and Information Retrieval* (STAIR), 5995766.
[20] M K M Nasution, S A M Noah, and S Saad 2011 Social network extraction: Superficial method and information retrieval *Proceeding of International Conference on Informatics for Development* (ICID'11), (arXiv:1601.02904v1 [cs.IR] 12 Jan 2016).
[21] K M N Mahyuddin, O S Sitompul, S Nasution, and H Ambarita 2017 New similarity *IOP Conference Series: Materials Science and Engineering* **180(1)**, 012297.
[22] M K M Nasution 2012 Simple search engine model: Adaptive properties *Cornell University Library* (arXiv:1212.3906v1 [cs.IR] 17 Dec 2012).
[23] M K M Nasution 2012 Simple search engine model: Adaptive properties for doubleton *Cornell University Library* (arXiv:1212.4702v1 [cs.IR] 19 Dec 2012).
[24] M K M Nasution 2013 Simple search engine model: Selective properties *Cornell University Library* (arXiv:1303.3964v1 [cs.IR] 16 Mar 2013).
[25] M K M Nasution and S A Noah 2012 Information retrieval model: A social network extraction perspective *Proceedings - 2012 International Conference on Information Retrieval and Knowledge Management*, (CAMP'12), 6204999.