Social network extraction based on Web: 4. A framework

Mahyuddin K M Nasution\textsuperscript{1,*}, Shahrul Azman Mohd Noah\textsuperscript{2}
\textsuperscript{1}Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Kampus USU, Padang Bulan 20155, Medan, Indonesia
\textsuperscript{2}Knowledge Technology Group Research, Center for Artificial Intelligence Technology, FTSM, Universiti Kebangsaan Malaysia, Bangi, Selangor DE, Malaysia.
E-mail: \textsuperscript{*}mahyuddin@usu.ac.id

Abstract. Social network extraction from information sources, as a research, requires the right steps. However, it is not enough, it need accurate description that explains the steps in one framework so that the flow of the extraction implementation information is well, as well as the extraction occurs systematically. In this paper, a framework for social network extraction has been outlined by involving several formulations and measurements as an explanation.

1. Introduction
At present, the entity that is most needed in decision making related to anything \cite{1, 2}, for example, from valuation to evaluation \cite{3}, from determining policies, decisions and penalties \cite{4}, or determining its other interests is the attendance of trusted relationships between entities \cite{5}, such as correlations in regression \cite{6}, or the equivalence principle in rough sets, etc \cite{7}. Therefore, in each entity that the role of humans as social beings is so important, the search for information on the relation between social actors becomes the focus of every life activity, whereby the entity itself becomes a social capital \cite{8}. Once the importance of that relationship information, every social networking system such as Facebook \cite{9}, Twitter \cite{10}, and others, tried to make an attraction \cite{11}, and is useful in determining business affairs on the behavior of social members \cite{12}. There are many relations between social networks that are created based on the causes of events \cite{12, 13, 14}. It requires the extraction of social networks, a task to crawl the information space so that the meaning behind the semantics of social activities can be recorded to accompany the social networking system.

2. Problem Definition
However, entities such as data, information, and knowledge, have a close relationship with one another \cite{15, 16, 17}. Although linearly, the relationship is through processing from one entity to another, but processing is an event that involves social actors, starting from those who collect data/information \cite{18}, those who encode / write it \cite{19}, those who study it \cite{20}, or those who presents it as information \cite{21}, and interpret it into knowledge \cite{22}. In reports or documents, recordings of all those forms go hand in hand with the name that records them \cite{23}. Then, that knowledge is used by them to determine the direction of their lives and influence others, or carry
out the social engineering. In this case, social actors are entities that may be individuals per person, individually, as a group or as an organization [24, 25, 26, 27, 28, 29].

A collection of social actors as entities and the relations among social actors as other entities form what is called as social networks [30]. Social network is a model of social structure with approaches to represent relations between individuals, groups, or organization [31]. Social networks have been shining for a long time since the term was introduced in 1954 [32], as a widely researched field of study. The formation of social networks involves data and information that is available or that has been successfully collected through records, historical track records, or through current events [31]. However, from the beginning until now, from individuals to organizations need data to build a strategy to achieve a goal [33]. Data processing functions to produce information that is useful in decision making. Of course, accurate data will return more effective results [34]. Forensically, such data depends on the person who gave it [35, 36].

In the information age, when computer networks connect social actors, it means social networks [37]. The study of social networks involving electronic documents both in the corpus or online has opened the door to new topics of social networking studies [38]. It breakthrough has brought new emphasis on social network research, and it has been stated as a policy in various related fields, especially in the field of information science or application of information technology [39], and then its application extends to other fields of study as well as the importance of data and information [40, 41, 42].

In general, the methods for extracting the social networks from information sources consist of two approaches [43, 44, 45, 46]. Information sources such as the corpus, consist of a collection of document $d_i, i = 1, \ldots, I$, or $D = \{d_i | i = 1, \ldots, I\}$ [47]. The corpus as a dataset, a collection of documents that have been specifically stated in terms of data processing interests, is in a semi-structured form. In contrast, by involving databases or online database such as DBLP, extracting social networks requires a special approach [31]. Databases are specifically structured data [48]. While information sources such as the Web contain unstructured data with the dynamics of changes to the content [49]. However, the web as a social media lacks semantic information and is not structured and should only be understood by humans [50]. Meanwhile, on the other hand the web is growing rapidly, becoming giant data that cannot be predicted [29]. Information about a social actor that has been recorded for a long time sometimes may still exist or has been excluded from the Internet environment. Only some information that is formally indexed, it remains part of the big data. Thus, the challenge of extracting social networks from information sources is not only related to the complexity of information, but also the limited achievements/access to that information [51].

3. Methodology

Social network extraction aims to retrieve relevant data from information sources based on data modelling that are in accordance with the provision of recording space [52], whereby data modelling to be a template of social network in this case. Suppose that $\Omega$ represents an information space where $|\Omega|$ is the size of $\Omega$, extracting social networks is as follows:

(i) Determining the existence of social actors. Social networks are formed from a number of social actors as main component of social network, it declares them as a collection of social actors, $A = \{a_i | i = 1, \ldots, n\}$. Suppose $\Omega_{a_i} \subseteq \Omega$ represents a social actor, if $|\Omega_{a_i}| > 0$, social actor $a_i$ is a part of social network. Or, $a_i \in \Omega$, i.e. the name of the social actor $a_i$ is already recorded in the information space.

(ii) Determining the relation of two social actors. Determining one of components for social networks is about relations between social actors. Identification of relation in the information space is based on the presence together in a document, a webpage, etc. For example $\Omega_{a_i} \subseteq \Omega$ and $\Omega_{a_j} \subseteq \Omega, i \neq j, j = 1, \ldots, n$, indicates the existence of two
social actors, then the existence of the relation between two social actors is stated with 
\[ |\Omega_{a_i} \cap \Omega_{a_j}| > 0 \] whereby \( \Omega_{a_i} \cap \Omega_{a_j} \subseteq \Omega \). Thus, the set of relations between two actors in
social networks is \( R = A \times A \), or \( \hat{R} = \{ r_k | k = 1, \ldots, K \} \).

Generally, a relation between two social actors in a social network has weight \( x_k \), making it possible to explore the meaning behind their relations. In the information space \( \Omega \) there is information that refers to relations, each information has its own weight of \( x_l \), \( l = 1, \ldots, L \), and the weight can be broken down into a collection of relations between two social actors, that is
\[ r(a_i, a_j)_l = x_k \times x_l, \quad r(a_i, a_j)_l \in R. \quad (1) \]

However, the information of social network that is present requires validation and this is recognized by information retrieval (IR) [53].

4. Framework
In general, formal social network extraction is to construct a graph \( G(V, E) \) where \( V = \emptyset \) is a set of vertices \( v_i, \ i = 1, \ldots, N \), whereas \( E \) is the set of edges \( e_k, \ k = 1, \ldots, K \) so there are \( \gamma_1 : A \xrightarrow{1:1} V \) and \( \gamma_2 : R \rightarrow E \) [54]. Thus, extracting social network from information sources requires two main steps, namely getting social actors and then adding relations among them [55]. However, in the first step there are obstacles where the source of information contains meaningful information [56]. Socially, the names of social actors have similarities in writing, such as "Mahyuddin Nasution" for different names: "Mahyuddin Nasution" (a lecturer) and "Muhammad Nasution", besides the same name: "Mahyuddin Nasution" (as a member of parliament). Thus, to ensure the relevance of information to a social actor, it is necessary to do a name disambiguation [57]. On several occasions, the implementation involves keywords as a companion to reduce information bias [24].

Thus, the extraction of social networks from information sources in the framework description is as follows [1]:

(i) Matching [55, 58]: This principle involves the presence together in a document or what is known as co-occurrence. Suppose the \( \Omega \) information space contains web pages as sources of information, or \( \omega_j \in \Omega, \ j = 1, \ldots, J \), if there is information for two social actors like \( \Omega_{a_i}, \Omega_{a_j} \subseteq \Omega \), so this principle presents
\[ p_r = |\Omega_{a_i} \cap \Omega_{a_j}| \quad (2) \]

The existence of a relation between two social actors is based on Eq. (2), that is, with a guarantee that it is valid if \( p_r > 0 \), whereby \( \Omega_{a_i} \cap \Omega_{a_j} \subseteq \Omega \), and \( p_r \) is expressed as the matching coefficient. The concept of matching also influences the information present from the information space, by considering that one of the social actors is the key to the existence of another [59]. There are cases like:

(a) Case 1.1: Suppose \( a_1 = Mahyuddin K. M. Nasution \) and \( a_2 = Opim Salim Sitompul \), with the query \( q = Mahyuddin K. M. Nasution, Opim Salim Sitompul \) submitted to the search engine, Google for example, and at the time of this writing it generated \( p_r = 1,150 > 0 \). The results of the search engine returns for this query are called hit count.

(b) Case 1.2: Another approach to the concept of matching is to consider the matching pattern of the name, namely \( a_1 = "Mahyuddin K. M. Nasution" \) and \( a_2 = "Opim Salim Sitompul" \) whereby query \( q = "Mahyuddin K. M. Nasution", "Opim Salim Sitompul" \). If this query is submitted to the same search engine, there is hit count \( p_t = 1,040 > 0 \).

(c) Case 1.3: Involving keywords like \( t_x = Universitas Sumatera Utara \) for example, gives a query in the form of \( q = Mahyuddin K. M. Nasution, Opim Salim Sitompul, Universitas Sumatera Utara \). A search engine, Google for example, generates hit count \( p_r = 1,110 > 0 \).
Table 1. Hit count as a result of returning Google search engine for different queries.

| Cases | \(|\Omega_{a_i} \cap \Omega_{a_j}|\) | Cases | \(A\) | \(|\Omega_{a_i}|\) |
|-------|----------------|-------|-----|-------|
| 1.1   | 1,150 \(\leq\) 2.1 | \(a_1\) | 46,000 |
| 1.1   | 1,150 \(\leq\) 2.1 | \(a_2\) | 6,950  |
| 1.2   | 1,040 \(\leq\) 2.2 | \(a_1\) | 5,640  |
| 1.2   | 1,040 \(\leq\) 2.2 | \(a_2\) | 6,620  |
| 1.3   | 1,110 \(\leq\) 2.3 | \(a_1\) | 29,800 |
| 1.3   | 1,110 \(\leq\) 2.3 | \(a_2\) | 11,900 |
| 1.4   | 472 \(\leq\) 2.4 | \(a_1\) | 1,800  |
| 1.4   | 472 \(\leq\) 2.4 | \(a_2\) | 3,860  |

(d) **Case 1.4:** Combination of the keyword involvement \(t_x\) and the pattern for which the query is in a form: \(q = "Mahyuddin K. M. Nasution", "Opim Salim Sitompul", "Universitas Sumatera Utara". The same search engine returns hit count for this query as follows: \(p_r = 472 > 0\).

The assessment of search engine returns, as the cases above have shown, is true that \(p_r(Case 1.1) \geq p_r(Case 1.3)\) and \(p_r(Case 2.1) \geq p_r(Case 1.4)\). Although, **Case 1.3 and Case 1.4** provide a different picture of the matching principle, that is

\[
q_r = |\Omega_{a_1} \cap \Omega_{a_j} \cap \Omega_{t_x}|
\]

whereby \(\Omega_{a_1} \cap \Omega_{a_j} \cap \Omega_{t_x} \subset \Omega [60, 61]\).

(ii) **Existence** [62, 63, 64]: Reveal the existence of social actors in the information space by proving that \(|\Omega_{a_1}| > 0\) for \(a_1 \in A, i = 1, \ldots, N [54]\).

The application of this concept to the query and search engine engagements contains the following:

(a) **Case 2.1:** Suppose \(a_1 = Mahyuddin K. M. Nasution\) and \(a_2 = Opim Salim Sitompul\) in queries form are \(q_{a_1} = Mahyuddin K. M. Nasution\) and \(q_{a_2} = Opim Salim Sitompul\). When those queries are submitted separately to the search engine, Google for example, returns hit count as a result each as follows \(q_{a_1} = 46,000 > 0\) and \(q_{a_2} = 6,950 > 0\).

(b) **Case 2.2:** While considering the matching pattern of the name, i.e. \(a_1 = "Mahyuddin K. M. Nasution"\) and \(a_2 = "Opim Salim Sitompul"\), the search engine returns hit count as a result respectively for \(a_1\) and \(a_2\) as is as follows \(q_{a_1} = 5,640 > 0\) and \(q_{a_2} = 6,620 > 0\).

(c) **Case 2.3:** Involving keywords like \(t_x = Universitas Sumatera Utara\) for example, gives a separate query in the form of \(q_{a_1} = Mahyuddin K. M. Nasution, Universitas Sumatera Utara\) and \(q_{a_2} = Opim Salim Sitompul, Universitas Sumatera Utara\). Furthermore, the search engine, Google for example, generates hit count as a result of the queries each as follows \(q_{a_1} = 29,800 > 0\) and \(q_{a_2} = 11,900 > 0\).

(d) **Case 2.4:** Combination of keyword \(t_x\) and two name patterns forms the different queries, those queries separately have form as follows \(q_{a_1} = "Mahyuddin K. M. Nasution", "Universitas Sumatera Utara"\) and \(q_{a_2} = "Opim Salim Sitompul", "Universitas Sumatera Utara"\). The same search engine generates hit count for them as results, where separately from each query is \(q_{a_1} = 1,800 > 0\) or \(q_{a_2} = 3,860 > 0\).

Thus, information justification for every social actor occurs when it involves a name pattern, and it guarantees the existence of social actors in the information space [65].

(iii) **Validation:** The implementation of the endorsement aims to guarantee the correctness of the results of the information return for each social actor and information about relations between two social actors. The information is said to be valid if it fulfills

\[
|\Omega_{a_1} \cap \Omega_{a_j}| \leq |\Omega_{a_1}|
\]
Table 2. Strength relation between two social actors

| Kasus | $\Omega_a_1$ | $\Omega_a_2$ | $\Omega_a_1 \cap \Omega_a_2$ | $J_c$ |
|-------|-------------|-------------|-----------------------------|------|
| 1     | 46,000      | 6,950       | 1,150                       | 0.0222 |
| 2     | 5,640       | 6,620       | 1,040                       | 0.0927 |
| 3     | 29,800      | 11,900      | 1,110                       | 0.0273 |
| 4     | 1,800       | 3,860       | 472                         | 0.0909 |

Table 3. Recall and precision for two cases

| Pendekatan | $D_k$       | $Ket$          |
|------------|-------------|----------------|
| Kasus 1    | 30/134 (22.30)% | 30/348 (8.62)% |
| Kasus 2    | 80/134 (59.70)% | 80/630 (12.70)% |

and

\[|\Omega_a_i \cap \Omega_a_j| \leq |\Omega_a_j|\]  \hspace{1cm} (5)

Implementation Eqs. (4) and (5) to all cases has the summary as shown in Table 1. In other words, if Eqs. (4) and (5) are not in accordance with the results in Table 1, the information returned by the search engine contains bias, which also causes the social networks do not have reliable information [66].

(iv) Strength relation: In general, strength in relationships is based on measurement of similarity. The similarity that is commonly used and has symmetrical properties in the measurement is the Jaccard coefficient

\[J_c = \frac{|\Omega_a_i \cap \Omega_a_j|}{|\Omega_a_i \cup \Omega_a_j|} = \frac{|\Omega_a_i \cap \Omega_a_j|}{|\Omega_a_i| + |\Omega_a_j| - |\Omega_a_i \cap \Omega_a_j|}\]  \hspace{1cm} (6)

where $J_c \in [0, 1]$, see Table 2.

(v) Survey: Preparing a dataset as a comparison is necessary by conducting a survey, either through a questionnaire of suitable social actors or gathering information related to social actors and their relationship with others by experts [54, 59, 60].

(vi) Information Retrieval: Let $E_R$ is the relevant set of edges existing social network (dataset), and $E_C$ is the achieved set of edges. The formula for calculating recall ($Rec$) and precision ($Prec$) is as follows [67]:

\[Rec = \frac{|E_R \cap E_C|}{|E_R|}\]  \hspace{1cm} (7)

and

\[Prec = \frac{|E_R \cap E_C|}{|E_C|}\]  \hspace{1cm} (8)

Table 3 explains the implementation of Eqs. (7) and (8).

5. Conclusion

The framework has provided a description brief of the extraction of social networks from information sources, related to information, data and structure, validation and assessment, and some of the requirements in the implementation of social network extraction.
References

[1] M K M Nasution 2010 Kerangka deskriptif untuk ekstraksi jaringan sosial Translasi, 1. DOI:10.13140/RG.2.2.24035.91688.

[2] M Janssen, H van der Voort, A Wahyudi 2017 Factors influencing big data decision-making quality Journal of Business Research 70. DOI:10.1016/j.jbusres.2016.08.007.

[3] M Munasinghe 1992 Environmental economics and valuation in development decision making Working Paper. World Bank, Environment Department, Washington, D.C., USA.

[4] J W Forrester 1992 Policies, decisions and information sources for modeling European Journal of Operational Research 59(1). DOI:10.1016/0377-2217(92)9006-U.

[5] J Wu, F Chidana 2014 A social network analysis trustconsensus based approach to group decision-making problems with interval-valued fuzzy reciprocal preference relations Knowledge-Based Systems 59. DOI:10.1016/j.knosys.2014.01.017.

[6] L C Briand, B Freimut, F Vollei 2004 Using multiple adaptive regression splines to support decision making in code inspections Journal of Systems and Software 73(2).

[7] C Yoe 2019 Principles of Risk Analysis Boca Raton - CRC Press. DOI:10.1201/9780429021121.

[8] S P Borgatti, C Jones 1998 Network measures of social capital Connections 21(2).

[9] N B Ellison, C Steinfield, C Lampe 2007 The benefits of Facebook "Friends:" Social capital and college students’ use of online social networks sites Journal of Computer-Mediated Communication 12(4). DOI:10.1111/j.1083-6101.2007.00367.x

[10] S A Myers, A Sharma, P Gupta, J Lin 2014 Information network or social network?: the structure of the twitter follow graph WWW '14 Companion Proceedings of the 23rd International Conference on World Wide Web. DOI:10.1145/2567948.2576939.

[11] Y Duan, J S Edwards, Y K Dwivedi 2019 Artificial intelligence for decision making in the era of Big Data - evolution, challenges and research agenda International Journal of Information Management bf 48. DOI:10.1016/j.ijinfomgt.2019.01.021

[12] R S Burt 2001 Attachment, decay, and social network Journal of Organizational Behavior 22(6). DOI:10.1002/job.106.

[13] A K Mackworth, E C Freuder 1985 The complexity of some polynomial network consistency algorithms for constraint satisfaction problems Artificial Intelligence 25(1). DOI:10.1016/0004-3702(85)90041-4.

[14] W Tan, M B Blake, I Saleh, S Dustdar 2013 Social-network-sourced big data analytics IEEE Internet Computing 17. DOI:10.1109/MIC.2013.100.

[15] C Zins 2007 Conceptual approaches for defining data, information, and knowledge Journal of the American Society for Information Science and Technology 58(4). DOI:10.1002/asi.20508.

[16] M Chen, D Ebert, H Hagen, R S Lamaree, K-L Ma, W Ribarsky, S Dustdar, D Silver 2008 Data, information, and knowledge in visualization IEEE Computer Graphics and Applications 29(1). DOI:10.1109/MCG.2009.6.

[17] M Boisot, A Canals 2004 Data, information and knowledge: have we got it right? Journal of Evolutionary Finance 14(1). DOI:10.1007/s00191-003-0181-9.

[18] S N Pal, C Duncombe, D Falzon, S Olsson 2013 Who Strategy for collecting safety data in public health programmes: complementing spontaneous reporting systems Drug Safety 36(2). DOI:10.1007/s40264-012-0014-6.

[19] T Sugahara, M Nishitani, W Inoha 2003 Variable length data encoding method, variable length data encoding apparatus, variable length encoded data decoding method, and variable length encoded data decoding apparatus US Patent App. 10/720,542.

[20] M K M Nasution, I Aulia, M Elveny 2019 Data Journal of Physics: Conference Series 1235(1). DOI:10.1088/1742-6596/1235/1/012110.

[21] L L Pipino, Y W Lee, R Y Yang 2002 Data quality assessment Communications of the ACM 45(4ve).

[22] J Woolnough 2000 How do students learn to apply their mathematical knowledge to interpret graphs in physics? Research in Science Education 30(3). DOI:10.1007/BF02461633.

[23] M K M Nasution 2018 Semantic interpretation of search engine resultant documents Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8397 LN.AI(PART 1). DOI:10.1007/978-3-319-05476-6-9.

[24] A McCallum 2005 Information extraction: distilling structured data from unstructured text Magazine Queue - Social Computing 3(9). DOI:10.1145/1105664.1105679.

[25] R M Alqulil, R M Alqulil, I Y Alskharova 2016 Extraction of hidden social networks from wiki-environment involved in information conflict International Journal of Intelligent Systems and Applications 8(2). DOI:10.5815/ijissa.2016.02.03.
[27] T Mooradian, B Renzi, K Matzler 2006 Who Trusts? Personality, Trust and Knowledge Sharing Management Learning 37(4). DOI:10.1177/1350507606073424.

[28] M K M Nasution, O S Sitompul 2017 Enhancing extraction method for aggregating strength relation between social actors Advances in Intelligent Systems and Computing 573. DOI:10.1007/978-3-319-57261-1_31.

[29] M K M Nasution, M Elveny, R Syah, S A Noah 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, IICEEI. DOI:10.1109/IICEEI.2015.7352551.

[30] M K M Nasution, S A Noah 2010 Superficial method for extracting social network for academics using web snippets Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6401 LNCS.

[31] M K M Nasution, S A Noah 2011 Extraction of academic social network from online database 2011 International Conference on Semantic Technology and Information Retrieval, STAIR 2011. DOI:10.1109/STAIR.2011.5995766.

[32] E F Churchill, C A Halverson 2005 Guest Editors’ Introduction: Social networks and social networking IEEE Internet Computing. DOI: 10.1007/978-3-642-16248-0_68.

[33] M K M Nasution, S A Noah 2017 Social Network Extraction Based on Web. A Comparison of Superficial Methods Procedia Computer Science 124. DOI:10.1016/j.procs.2017.12.133.

[34] M K M Nasution 2019 Information retrieval optimization based on tree of social network IOP Conference Series: Materials Science and Engineering 505(1). DOI:10.1088/1757-899X/505/1/012009.

[35] 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, CIITSM 2016. DOI:10.1109/CITSM.2016.7577513.

[36] M K M Nasution, M Hardi, R Sitepu, E Sinulingga 2017 A Method to extract the forensic about negative issues from Web IOP Conference Series: Materials Science and Engineering 180(1). DOI:10.1088/1757-899X/180/1/012241.

[37] B Wellman 2001 Computer networks as social networks Science 293(5537). DOI:10.1126/science.1065547.

[38] H Kautz, B Selman, A Miliewski 1996 Agent amplified communication Proceedings of AAAI-96.

[39] S Jain, A Sinha 2019 Social network sustainability for transport planning with complex interconnections Sustainable Computing: Informatics and Systems 24. DOI:10.1016/j.suscom.2019.100351.

[40] S J Polizzi, B Ofem, W Coyle, K Lundquist, G T Rushton 2019 Social network data from teacher leader Journal of Systems and Software 14(2). DOI: 10.1017/mor.2018.13.

[41] T Tsekeris, C Tsekeris, I Katerelos 2018 Reflections on networks, human behaviour, and social dynamics in the digital age AI and Society 33(2). DOI: 10.1007/s00146-017-0704-9.

[42] M Elveny, R Syah, M Elfida, M K M Nasution 2018 Information Retrieval on social network: An method for extracting Social network by the relation existence Advances in Intelligent Systems and Computing 573. DOI:10.1007/978-3-319-57261-1_31.

[43] M Elveny, R Syah, M Elfida, M K M Nasution 2018 Information Retrieval on social network: An method for extracting Social network by the relation existence Advances in Intelligent Systems and Computing 573. DOI:10.1007/978-3-319-57261-1_31.

[44] J Dietrich, N Jones 2008 Using social networking and Semantic Web technology in software engineering - use case, patterns, and a case study Journal of Systems and Software 81(12). DOI:10.1016/j.jss.2008.03.060.

[45] Y Matsuo, J Mori, M Hamasaki, T Nishimura, H Takeda, K Hasida 2006 POLYPHONET: An advanced social networks extraction system Proceedings of the 15th International Conference on World Wide Web 2006 (WWW 2006).

[46] M Elveny, R Syah, M Elfida, M K M Nasution 2018 Information Retrieval on social network: An method for extracting Social network by the relation existence Advances in Intelligent Systems and Computing 573. DOI:10.1007/978-3-319-57261-1_31.
[53] P Mika 2007 *Social Networks and the Semantic Web* Springer-Verlag, Berlin.

[54] M K M Nasution 2018 Social network extraction based on Web: 1. Related superficial methods *IOP Conference Series: Materials Science and Engineering* **300**(1). DOI:10.1088/1757-899X/300/1/012056.

[55] Y. Matsuo, J Mori, M Hamasaki, T Nishimura, T Takeda, K Hasida, M Ishizuka 2007 POLYPHONET: An advanced social networks extraction system from the Web *Web Semantics* **5**(4). DOI:10.1016/j.websem.2007.09.002.

[56] R Bekkerman, A McCallum 2005 Disambiguating Web appearances of people in a social network *Proc. WWW 2005*. DOI:10.1145/1060745.1060813.

[57] M K M Nasution 2015 Extracting Keyword for Disambiguating Name Based on the Overlap Principle *Proceedings of International Conference on Information Technology and Engineering Application* (4-th ICIBA), *Book 1*. arXiv:1602.00104v1 [cs.IR].

[58] M K M Nasution, S A Noah 2012 A Methodology to Extract Social Network from the Web Snippet *Cornell University*. arXiv:1211.5877v1 [cs.SI].

[59] M K M Nasution 2018 Social network extraction based on web: 2. Strategies in superficial methods *Journal of Physics: Conference Series* **1116**(2). DOI:10.1088/1742-6596/1116/2/022029.

[60] M K M Nasution, O S Sitompul, S A Noah 2018 *Journal of Physics: Conference Series* **978**(1). DOI:10.1088/1742-6596/978/1/012033.

[61] M K M Nasution 2018 Doubleton: A Role of the Search Engine to Reveal the Existence of Relation in Information Space *IOP Conference Series: Materials Science and Engineering* **420**(1). DOI:10.1088/1757-899X/420/1/012138.

[62] D Bollegala, Y Matsuo, M Ishizuka 2006 Extracting key phrase to disambiguate personal names on the Web *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **3878 LNCS**. DOI:10.1007/11671299_24.

[63] D A Pereira, B Ribeiro-Neto, N Ziviani, A H F Laender, M A Gonalves, A A Ferreira 2009 Using web information for author name disambiguation Proceedings of the ACM/IEEE Joint Conference on Digital Libraries. DOI: 10.1145/1555400.1555409.

[64] I Hussain, S Asghar 2018 Resolving namesakes using the author’s social network *Turkish Journal of Electrical Engineering and Computer Sciences* **26**(1). DOI:10.3906/elk-1702-293.

[65] M K M Nasution 2018 Singleton: A Role of the Search Engine to Reveal the Existence of Something in Information Space *IOP Conference Series: Materials Science and Engineering* **420**(1). DOI:10.1088/1757-899X/420/1/012137.

[66] S Dias, A J Sutton, N J Welton 2013 Evidence Synthesis for decision making 3: Heterogeneity subgroups, meta-Regression, bias, and bias-adjustment *Medical Decision Making* **33**(5). DOI:10.1177/0272989X13485157.

[67] M K M Nasution 2013 Superficial method for extracting academic social network from the Web *Ph.D Thesis* Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia.