

Integrated Social Media Knowledge Capture in Medical Domain of Indonesia

Kridanto Surendro*, Dicky Prima Satya, Farrell Yodihartomo

School of Electrical Engineering and Informatics, Institut Teknologi Bandung (ITB)

Jalan Ganesha 10 Bandung, Phone: +62-22-2508135, Indonesia

*Corresponding author, e-mail: endro@informatika.org

Abstract

The Social Media Platforms, as the one of largest part of today data traffic on the Internet, disseminate a vast volume of information, including medical information in it. Knowledge management system (KMS) approach is applied with purpose to capture, maintain, and manage tacit or explicit knowledge available and collected within the social media platforms, organization's database, knowledge base, or document repository. By adding Indonesian Natural Language Processing (InaNLP), Machine Learning and Data Mining approach, our research has proposed a framework which is theoretically designed to improve the previous research related to social media knowledge capture model and enhance its accuracy and reliability of knowledge retrieved compared to previous knowledge capture model. This system mainly aimed for medical practitioner to give a quick suggestion of the diseases regarding to the early diagnose which has been taken in the first place. On this current research state, the pre-processing phase of the framework implementation and knowledge presentation is our main concern to maximize the information value for the knowledge users and also to reduce the language issues in texts such as ambiguity, inconsistency, use of slang vocabulary, etc. According to this research's goal, we have designed an algorithm to extract feature from dataset.

Keywords: Knowledge management system, Natural language processing, Data mining, Machine learning, Medical knowledge

Copyright © 2018 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

Healthcare system in developing country nowadays are having a difficult state to accommodate health issues that occur within the country, especially in Indonesia. Despite the internet access is growing in big cities, people still have to make efforts to find a reliable source of knowledge of medical, in this context are drugs and medical actions should be taken. The knowledge become useful when it is maintained and can be accessible for those who need it to cure disease. Since the medical sector have become increasingly important today, knowledge creation and diffusion is necessary to cater in giving better health services and quick-response first aid actions.

Social media distributes a highly diverse of information through the internet, including text, image, video, etc. A vast number of media contents which are disseminated through the social media contain any kind of useful information that can be used in order to gain knowledge [1]. By using the NLP and machine learning approach, it becomes possible to extract the specific knowledge for supporting the credibility of KMS designed.

Previous research in terms of Knowledge Capture System for Medical Domain has shown the prototype can predict the relevant knowledge which retrieved by the user with only 38.89% accuracy using 62 records of training dataset. This result caused due to its limitations: lack of collected data for creating the predictive model; and the knowledge capture model pre-processed the collected model manually [2]. According to this research's goal, we have proposed a framework that could improve the previous knowledge capture model accuracy and reliability in the model creation, preprocessing, and knowledge sharing phase [3]. Since our framework is generally designed based on literature review, it becomes our concern to set goals for the development and evaluation of the framework to be conducted in our works.

We try to develop the Web-based KMS according to this framework in pursuance of framework evaluation result whether the framework is actually able to improve the result of

previous study. The pre-processing phase of the model implementation and knowledge presentation is our main concern on this current research state. We have designed an algorithm to extract feature from dataset to maximize the information value for the knowledge users and also to reduce the language issues in texts such as ambiguity, inconsistency, use of slang vocabulary, etc.

In the future, information technologies will be obiquitously applied in healthcare and medical organizations in order to meet the medical requirements with proper health services, suggestions, or decision making. In addition, the use of information technology (IT) in medical domain enables a holistic management of medical knowledge and its secure exchange between healthcare or medical providers and recipients whom may receive its advantages [4]. By integrating these approach, it creates a greater range of possibility to improve medical services, especially in Indonesia.

2. Literature Review

2.1. Perspective on Knowledge, Knowledge Management and Knowledge Capture

In many years, knowledge has been produced and used by human to improve and enhance their value of life. Despite, the usefulness of knowledge itself is relative to humans understanding, insight, and how knowledge can be used in practice [5]. Since knowledge deliver values, people conduct knowledge exchange in many ways, in example discussion of two people through a deep conversation and it is documented both formally and informally [6].

According to Ackoff's idea on DIKW (Data, Information, Knowledge, Wisdom) hierarchy [7], knowledge can be generated when the body of information is organized in a particular way and connected each other in order to deliver message or fact. This knowledge generation may be executed by individual or group [8]. In knowledge management framework context, four steps of Wiig's KM cycle as shown on Figure 1 represent how knowledge is created and used as individuals or as groups [5, 8]. This also relates to the process how the knowledge is being captured.

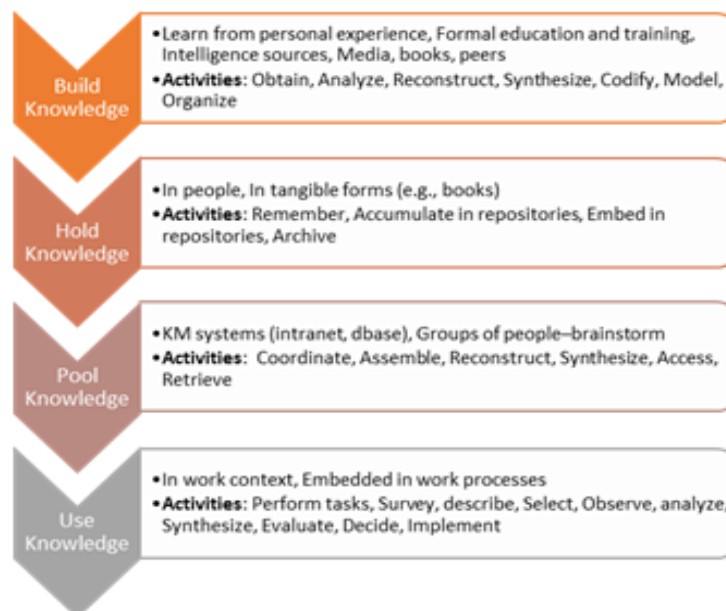


Figure 1. Wiig's KM Cycle Major Steps [4]

Greenes Consulting [9] declare 10-step guide to knowledge capture which helps codifying the knowledge process. This makes possible to enhance and measure the KM process by associating the 10-step guide to knowledge capture for structuring, managing, and

disseminating knowledge [10–12]. The 10-step guide to knowledge capture is described as follows [9]:

1. Identify customer for the knowledge
2. Identify a community of practice
3. Get clear what the captured knowledge is really about
4. Collate any existing material of captured knowledge
5. Elicit knowledge
6. Distil the knowledge
7. Organize and package the knowledge with the customer in mind
8. Validate the guideline
9. Publish the knowledge
10. Initiate a feedback and ownership process

According to the following steps, it is possible to enhance the KM process by associating the 10-step guide to knowledge capture for a better understanding of knowledge management.

2.2. Natural Language Processing for Indonesian Language

Nowadays, NLP applications depend on various pre-processing steps, such as Chunking, Tokenization, Stemming, Lemmatization, POS tagging, sentence boundary identification, Named Entity Recognition (NER), etc. [13]. However, methods used in the NLP for text mining have limitations in the matter of ambiguity and inconsistency within sentences.

Indonesian language (Bahasa Indonesia) has its own characteristic due to the region influences and historical vocabulary development. Bahasa Indonesia has an indulgent sentence structure, particularly in daily conversation both in a form of text or speech communication. Although the word order within a phrase is changed or some of the words are missing, the sentence is still can be understood roughly [14]. An architecture has been developed to solve this problem. InaNLP System Architecture as shown in Figure 2 is developed with concerning the sentence structure of Bahasa Indonesia. InaNLP contains a number of NLP module where each module can be flexibly applied and added as a dependency by the user [14].

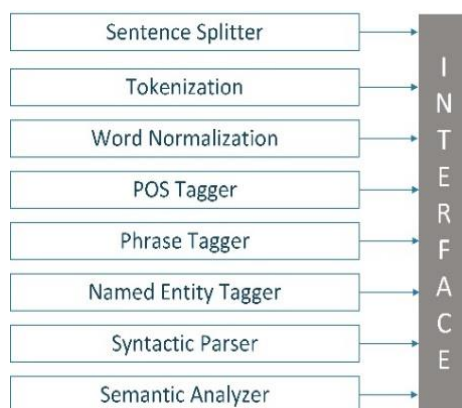


Figure 2. InaNLP Architecture [14]

2.3. Machine Learning and Data Mining

Machine Learning and Data mining has become an important research area since the market of database industry needs of extracting valuable knowledge from data sources is constantly growing. Extracting implicit and potential data is a part of knowledge discovery process which become a generic goal for industry [15]. In order to cater this goal, various of machine learning algorithm has been developed such as decision trees, rule-based, bayesian network, nearest-neighbour, nonlinear regression, etc [16]. Combining the data analytical needs and modeling tools, several data mining frameworks have been developed, in example CRISP-DM (Cross-industry Standard Process for Data Mining) [17]. This framework is generic and it is able to carry out data mining project in many industry sectors.

2.4. Social Media in Healthcare & Medical Sector

Recently, there are online doctor communities among professional healthcare and medical on social networks which exchange information on such subject related to clinical diagnosis, bodily symptoms, treatment options, experience of medical evidence, and their opinion about the medical treatment [18]. Moreover, 29% of patients looking for health information through social media and viewing others experiences with particular disease [19]. As a result of that, patients are more independent and well prepared when they are equipped with medical information and this leads to a greater influence of patients' decision making [20,21].

2.5. Existing Work-social Media Knowledge Capture Model. Case Study: Medical Knowledge

Cestyakara [2] proposed a knowledge capture model, as shown in Figure 3, for optimizing knowledge management process. This model implements KMS and data mining concepts. These activities in this framework are derived from the 10-step Guide to Knowledge Capture [9].

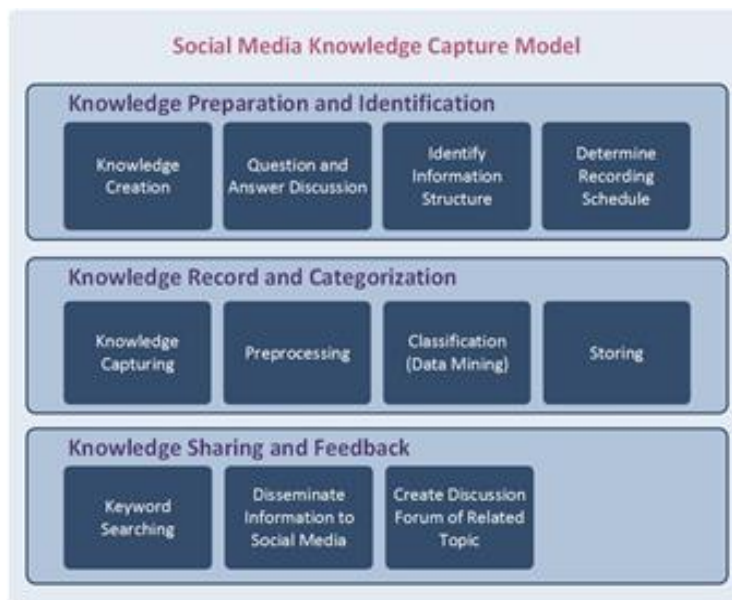


Figure 3. Cestyakara's Social Media Knowledge Capture Model [2]

Knowledge source of this conducted research is retrieved from the social media through the Facebook Graph API which contains of medical information within forum discussions. In the pre-processing step, the texts which gathered from the social media is processed using manual approach to extract the features and build a training corpus. Using WEKA [22] as a tool for classification learning, this model produces predicted categories based on 11 medical features (e.g. tips, treatment, disease, symptom, etc.).

Feedbacks and effectiveness of the predicted knowledge are tested through the expert judgment and online forum/fanpage discussion. According to this model, there are opportunities to improve the accuracy of the existing model and add automations on pre-processing the texts retrieved. This research is conducted to continue the improvement of KMS model by adopting relevant approaches.

3. Integrated Social Media Knowledge Capture

Respecting to the methodology and approach used in this research, a framework is proposed to comprehend the optimization of the previous model. Our goal with the proposed

framework is to enable patient to capture hierarchical useful medical knowledge since the model act as assistive media to the patient.

3.1. Integrated Knowledge Capture Model

We evaluate Cestyakara's model [2] in each phases to gain the understanding and to know how far the model can elicit useful medical knowledge. This previous research proves that social media knowledge capture model can be implemented using existing technologies. As the result of this model implementation, there are some improvements required in the information pre-processing step by shifting the manual process into an automated by using an NLP technology, and prepare a larger dataset in order to achieve a more accurate training model.

Due to the evaluation of previous model, we establish a model which is theoretically possible to gain a better result. Since our proposed model accommodate end-to-end social media functionality and text processing improvement, several approaches are adopted to cater the research objective. According to the Wiig's KM Cycle [4], the proposed model follows the sequence of the knowledge management phase with associating the 10-step guide of knowledge capture on each section [19].

3.2. Integrated Social Media Knowledge Capture Framework

Since this model is a theoretical design, we combined concepts which has been explained in the Chapter 2 to get an end-to-end knowledge capture framework. There are four steps of this framework to be implemented as shown on Figure 4. First, in the Build Knowledge phase, source of knowledge is prepared through health and medical forums. In social network posts, comments, and tweets contain meaningful knowledge that can be captured through a natural conversations. Specific subject of conversation naturally emerged through the contents shared by those who shares their experience about their health condition whether it is a symptom, disease, or health problem. After identifying the content of conversation, it is possible to determine whether the information captured is structured or unstructured.

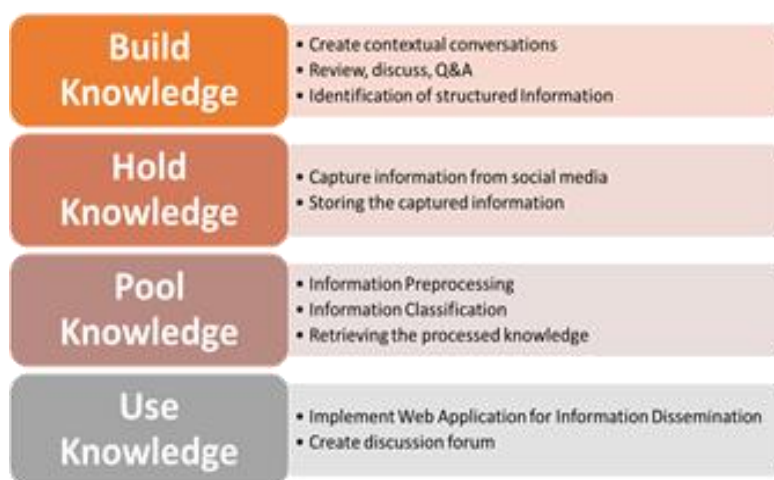


Figure 4. Integrated Knowledge Capture Framework

In the Hold Knowledge phase, captured information from the data gathered through the social media sources are filtered and transformed in purpose to gain meaningful sources of information needed which is used in generating a better quality of training model. Information retrieved is collected in a raw data form which has to be going through preprocessed step. This collected raw data saved in a DBMS (Database Management System) to make sure that raw data is stored and managed in a proper manner.

In the Pool Knowledge phase, as mentioned in the NLP section, inconsistency language used in the retrieved texts need to be formalized and pre-processed in the NLP module to extract medical features contained in it [14]. During this phase, InaNLP Framework modules are

activated for pre-processing the information gathered into a new form of pre-processed text dataset. The system model relies on lexical dictionary, word formalizer, set of NER, word tokenization, and POS Tagging rule, and training dataset from Bahasa Indonesia language in order to gain patterns of collected information in determining the medical feature. List of sentences are formalized in the first place before it is processed through the tokenizer and stemmer. Sentence splitter works in case where posts are retrieved in a paragraph form into a list of sentence. Each words and punctuation will be tokenized and stemmer work to transform tokenized word into its root word. After that, POS Tagger determined the Indonesian language tag for each word as an additive information. The features are determined information characteristic as default within the sentence.

Prepared with Medical Named Entity List as a corpus, this model will assign each instance of datasets with binary record within every instance for every determined features. We prepare a rule set to determine the binary set for each instance based on what kind of gazette that match with particular parsed word. Our model adopts the multi-label classifier to handle context of medical terms used or matched named entity for each instance to construct a dataset of binary record [23]. This because of the execution of Agastya's model determined a sentence with a single label while it possibly contains many of categories that would be classified and beneficial in other aspect of information [3]. In example, a patient can make a statement of his disease, action, body part, and medicine in a single sentence. We see that particular sentence has multiple information in it, then is should be handled in multi-label to maximize of capturing the information. In multi-label classifier, all set of features are able to be assigned as labels [23–25]. Therefore, the rule set must have a method to handle what kind of tag that should be marked as a label/predictive label beforehand. Multi-label dataset training allows predictive model to be produced and used for the predicting the input for the next captured information and evaluate the result. This predictive model and are stored in the file storage. After that, the result structured information is stored into a database management system to be accessed in the Use Knowledge phase.

In the Use Knowledge phase, the accuracy and retrieval of structured information is tested by developing a web 2.0 platform to evaluate the performance of conveyed medical knowledge capture is for patients or medical practitioners through information tags search on it and ensure the requirement of this system accommodates patient's expectation [26]. Furthermore, knowledge is evaluated not only through the predictive phase but also by asking for expert judgment.

3.3. Model Comparison

Since this framework is designed by combining various perspectives and existing model, it is important to examine in which aspect that this framework is excel or deprived. Therefore, we try to compare our framework to previous model [2] and CRISP-DM model [17] as shown in Table 1.

According to the Table 1, there are steps that nearly similar both in the Agastyas' model and our framework. Since our objectives is to improve the cycle, our model does not include the "Determine Record Schedule" activity. We define the condition of Build Knowledge step which is necessary to be conducted if the source of knowledge does not exist yet. Thus, if the knowledge does exist and is able to be retrieved from social media source, we can manage to collect data as much as required to build a dataset in order to improve model prediction accuracy with or without a schedule.

In our model, capturing information from social media and store it to a DBMS are distinct activities in particular KM Cycle Step with purpose to differentiate the understanding of the way we retrieved the data then hold it into a form that can be processed later. On the other hand, agastyas' previous model do the storing the data and the predictive model several steps after the phase where we obtain the knowledge. We take a preprocessing activity in agastyas' model into a more detail step. Since our model automate the preprocessing step using InaNLP module, constructing the dataset for predictive model training require additional steps in the back-end side. As explained in the previous section, our model consider the context of medical features to minimize the loss of information by using the single label predictive model. Thus, we apply the multi-label approach in our model [23, 25, 27].

We review the CRISP-DM Framework to obtain understanding of data mining process to enhance our machine learning steps. Our understanding to the CRISP-DM is mapped in the

Table 1. Since the CRISP-DM is designed for industrial sector and to provide analytical reports, it has a different purpose to our model. Our model is purposed to provide knowledge related to medical topic and relevant solution to meet particular medical issue. Due to this review, we realized in CRISP-DM approach does not apply the evaluate step in the end of cycle while our model require this step in the last phase of cycle to gain feedbacks from user and experts after they make use of the provided knowledge by the system which adopt our knowledge capture model [17].

Table 1. Knowledge Model Comparison

Integrated Knowledge Capture Framework				
KM Cycle Steps	Activity	Methodologies and Approaches	Agastyas' Knowledge Capture Model [2]	CRISP-DM [17]
Build Knowledge	Create contextual conversations	Knowledge Generation (on Discussion Media)	Create contextual conversations	Business Understanding
	Review, discuss, Q&A	Knowledge Generation (on Discussion Media)	Review, discuss, Q&A	Business Understanding
Hold Knowledge	Identification of structured Information	Knowledge Structure Understanding	Identification of structured Information	Data Understanding
	Capture information from social media	Data Source from Social Media Platform API	Knowledge Capturing	Data Preparation
	Storing the captured information	Database Management	Storing	Data Preparation
Pool Knowledge	Information Preprocessing	Cleaning, Transformation, and Attribute Selection	Preprocessing	Modeling
	Information Classification	Medical Context: Rule-based Dataset Creation and Feature Extraction	Preprocessing	Modeling
Use Knowledge	Retrieving the processed knowledge	Multi-label Classifier Algorithm and Predictive Model Generation	Classification (Data Mining)	Modeling, Evaluation
	Implement Web Application for Information Dissemination	Database Management	-	Modeling
	Create discussion forum	Develop (Web 2.0) Web-based KMS	Keyword Searching	Deployment
		Feedback and knowledge exchange media forum	Disseminate Information to Social Media, Create Discussion Forum of Related Topic	Evaluation

4. Integrated Knowledge Capture Framework Implementation on Indonesian Medical Knowledge Model

We implement the framework in pursuance of framework evaluation result whether the framework is actually able to improve the result of previous study. The implementation approaches for each activity of Integrated Knowledge Capture Framework is shown in Table 2.

Table 2. Implementation Approaches

KM Cycle Steps	Activity	Implementation Approaches
Build Knowledge	Create contextual conversations	Social Media in Healthcare and Medical Domain - Information exchange media
	Review, discuss, Q&A	Social Media in Healthcare and Medical Domain - Media repository for discussion history
Hold Knowledge	Identification of structured Information	Social Media in Healthcare and Medical Domain - To review the structure of information
	Capture information from social media	Data source from Facebook and Twitter API
Pool Knowledge	Information preprocessing, Text Mining, and Classification	InaNLP Framework (includes Machine Learning Algorithm for Syntactic Parser and POS Tagging) [19] CRISP - DM [15]
	Storing the processed knowledge	Database Management

Table 2. Implementation Approaches

KM Cycle Steps	Activity	Implementation Approaches
Use Knowledge	Implement Web Application for Information Dissemination	Develop (Web 2.0) Web-based KMS
	Create discussion forum	Feedback and knowledge exchange media forum

Detailed implementation of the framework on Indonesian medical knowledge model is shown on Figure 5. System capture information from social media such as Facebook and Twitter as representation of Social Media in Healthcare and Medical Domain information exchange media and media repository for discussion history, and store the result in a CSV file. These steps are the implementation of Build Knowledge and Hold Knowledge in KM Cycle Steps.

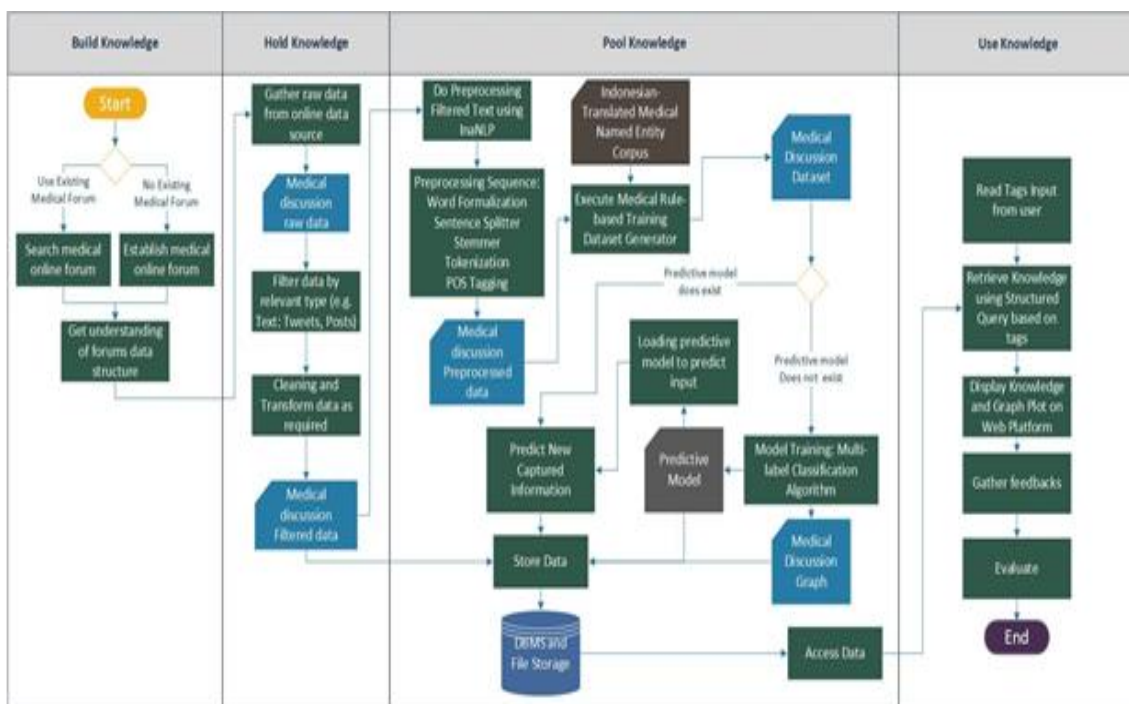


Figure 5. Integrated Knowledge Capture Framework Implementation on Indonesian Medical Knowledge Model

As representation of Pool Knowledge in KM Cycle Steps, the pre-processing phase of the model implementation and knowledge presentation is our main concern on this current research state. We are using InaNLP to reduce the language issues in texts such as ambiguity, inconsistency, use of slang vocabulary, informal sentences, etc. Example of preprocessing step of informal sentence which contain medical information using InaNLP is shown in Figure 6. We have designed an algorithm as shown on Figure 7 to extract feature from preprocessed dataset to maximize the information value for the knowledge users.

The algorithm will scan the file for certain attributes pattern for each social media information record. If a record has word that exist in word list of medical NER, system will write a flag in related attributes of medical NER. For example, one of medical NER is Disease. Disease has certain features/attributes like definition, medicine, prevention, indication, etc. After feature extraction, the next phase is system will check multi-label pattern which match one of existing rules. Rules will identify classification of social media information record and store it as processed knowledge complete with related medical NER and its feature/attributes. As representation of Use Knowledge, the last phase in KM Cycle Steps, our future work is implementation Web Based KM for processed knowledge dissemination.


```

original: "pagi, dok. sy Budiharjo. ud 2 mlm sy sakit pinggang ngilu-
ngilu n terasa beradh. sy hrs makan apa ya dok?"
data:
  0:
    sentence: "pagi, dok ."
    formalized: "pagi , dok ."
    token:
      0: "pagi"
      1: ","
      2: "dok"
      3: "."
    stem:
      0: "pagi"
      1: ","
      2: "dok"
      3: "."
    postag:
      0: "NN"
      1: ","
      2: "NN"
      3: "."
    NE:
      0: "pagi OTHER"
      1: ", OTHER"
      2: "dok OTHER"
      3: ". OTHER"
  1:
    sentence: "sy Budiharjo ."
    formalized: "saya Budiharjo ."
    token: [-]
    stem: [-]
    postag: [-]
    NE: [-]
  2: (-)
  3: (-)

```

Figure 6. Preprocessing step of informal sentence which contain medical information using InaNLP

Dataset Feature Extraction Algorithm

```

#adding dependencies
require medical_discussion_posts_file, translated_INA_MeSH_file, InaNLP_NER_file, csv_multilabel_dataset,
medical_rule_set
require BMH_algorithm

#instantiation
medical_posts : load (medical_discussion_posts_file)
medical_NER : load (translated_INA_MeSH_file)
generic_NER : load (InaNLP_NER_file)
csv_file : load (csv_multilabel_dataset)
rule_set : load (medical_rule_set)

#algorithm

#add new datarecord/
set read pointer csv_file.record to where the record is null
foreach medical_posts.posts in medical_posts
  write medical_posts.posts to the first column of csv_file.record

#feature extraction
##-at this phase, csv file only contains the attribute pattern for each record
foreach csv_file.record in csv_file
  check if csv_file.record has empty attribute and class is true then
  foreach word in csv_file.record and not end of medical posts file
    do BMH_algorithm to check if word exist in medical_NER, generic_NER
    if word found then
      get NER attribute of (medical_NER or generic_NER)
      write '1' to similar NER attribute in csv_file
    else write '0' to other unmatched attributes
  else go to the next csv_file.record
save (csv_file)

#Label extraction rule
##-at this phase, multi-Label/multi-class pattern will be added for each record
foreach record in csv_file
  do BMH_algorithm to check if csv_file.record pattern match to one of existing rule_set.rule
  if csv_file.record pattern match then
    write rule_set.class pattern to each csv_file record as an additional attribute
  else write rule_set.class default #default value is 0 for each distinct class
save (csv_file)

```

Figure 7. Dataset Feature Extraction Algorithm

5. Conclusion

Regarding to the healthcare and medical sector issues in Indonesia, many healthcare organizations is striving to increase the dependency of integrated knowledge management system so that they can maximize the use of medical knowledge and capture it for taking medical actions. It also brings a beneficial impact to the organization increasing their medical service quality and time efficiency.

In this research, we suggest an integrated approach which helps healthcare or medical organization to establish an end-to-end knowledge management system and enables stakeholders and medical practitioners to use and share knowledge in an effective way. We already create a framework of integrated knowledge capture. We also implemented the framework to build, hold, and pool knowledge. Although this framework theoretically improves several aspects of the previous research, yet it still has to be proven by implementing this framework into a prototype which can produce objective result. Our future work is to leverage the knowledge generated from research by building Web-based KMS for knowledge dissemination.

Acknowledgement

This study is financially supported by the Institut Teknologi Bandung (ITB) under Program Penelitian, Pengabdian kepada Masyarakat, dan Inovasi (P3MI).

References

- [1] S Tarik, M Kumar, SK Singh. *A Review of Social Media*. In Future. 2017; 8(3): 670–3.
- [2] A Cestyakara, K Surendro. *Social Media Knowledge Capture Model for Optimizing Knowledge Management Using Classification Learning Techniques*. Case Study: Medical Knowledge (Thesis - In Bahasa). Institut Teknologi Bandung; 2015.
- [3] K Surendro, F Yodihartomo, DP Satya. *Integrated Social Media Knowledge Capture Model in Medical Domain of Indonesia*.
- [4] J Gamble, PR. Blackwell. *Knowledge Management: A State of the Art Guide*. London: Kogan Page Ltd. 2001.
- [5] KM Wiig. *Knowledge Management: An Introduction and Perspective*. *Journal of Knowledge Management*. 1997; 1(1): 6–14.
- [6] F Bagayogo, L Lapointe, J Ramaprasad, I Vedel. *Co-Creation of Knowledge in Healthcare: A Study of Social Media Usage*. Proceedings of the Annual Hawaii International Conference on System Sciences. 2014; 626–35.
- [7] K Dalkir. *Knowledge Management in Theory and Practice*. Butterworth Heinemann. 2005; 4: 392.
- [8] K Dalkir. *Introduction to Knowledge Management*. *Knowledge Management in Theory and Practice*. 2001; 1–30.
- [9] Greenes Consulting. *10-Step Guide to Knowledge Capture*. 2010.
- [10] I Nonaka, R Toyama, P Byosièrè. *Theory of Organizational Knowledge Creation: Understanding the Dynamic Process of Creating Knowledge*. In: DIERKES, M.; et al. *Handbook Organizational Learning and Knowledge*. 2001: 491–517.
- [11] IS Suwardi, K Surendro, Aradea. *An Overview of Multi Agent System Approach in Knowledge Management Model*. International Conference on Information Technology Systems and Innovation, ICITSI 2014 - Proceeding. 2014; (November): 62–9.
- [12] LH Atrinawati, K Surendro. *Assessment for Knowledge Management Readiness*. Proceedings of the 2009 International Conference on Electrical Engineering and Informatics, ICEEI 2009. 2009; 2(August): 399–404.
- [13] R Sunil, V Jayan, VK Bhadrans. *Preprocessors in NLP Applications: In the Context of English to Malayalam Machine Translation*. 2012 Annual IEEE India Conference, INDICON 2012. 2012; 221–6.
- [14] A Purwarianti, A Andhika, AF Wicaksono, I Afif, F Ferdian. *InaNLP: Indonesia Natural Language Processing Toolkit, Case Study: Complaint Tweet Classification*. 4th IGNITE Conference and 2016 International Conference on Advanced Informatics: Concepts, Theory and Application, ICAICTA 2016. 2016; 5–9.
- [15] J Fürnkranz, D Gamberger, N Lavrač. *Foundations of Rule Learning*. 2012; (2003).
- [16] U Fayyad, G Piatetsky-Shapiro, P Smyth. *Knowledge Discovery and Data Mining: Towards a Unifying Framework*. Int Conf on Knowledge Discovery and Data Mining. 1996; 82–88.
- [17] P Chapman, J Clinton, R Kerber, T Khabaza, T Reinartz, C Shearer, et al. *Crisp-Dm 1.0*. CRISP-DM Consortium. 2000; 76.
- [18] F Griffiths, J Cave, F Boardman, J Ren, T Pawlikowska, R Ball, et al. *Social Networks - The Future for Health Care Delivery*. *Social Science and Medicine*. 2012; 75(12): 2233–41.

-
- [19] Pricewaterhouse Coopers. Social Media “likes” Healthcare. *Health San Francisco*. 2012; (April): 40.
- [20] H Waitzkin, JD Stoeckle. *Information Control and the Micropolitics of Health Care*. In J. D. Stoeckle (Ed.), *Encounters between Patients and Doctors: An Anthology*. Cambridge Mass: MIT Press; 1976.
- [21] R Cullen. *Health Information on the Internet: A Study of Providers, Quality, and Users*. Prager. 2006: 256.
- [22] M Hall, E Frank, G Holmes, B Pfahringer, P Reutemann, IH Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explor Newsl*. 2009; 11(1): 10–8.
- [23] EA Tanaka, SR Nozawa, AA Macedo, JA Baranauskas. A Multi-Label Approach Using Binary Relevance and Decision Trees Applied to Functional Genomics. *Journal of Biomedical Informatics*. 2015; 54: 85–95.
- [24] G Started, D Format, M Datasets, U Meka, CL Interface, GU Interface, et al. *Tutorial*. Meka 1.9.1. 2017; (April): 1–19.
- [25] X Kong, B Cao, PS Yu. *Multi-Label Classification by Mining Label and Instance Correlations from Heterogeneous Information Networks*. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13. 2013; 614.
- [26] K Surendro, C Martini. Hierarchical I* Modeling in Requirement Engineering. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2016; 14(2): 784–90.
- [27] G Tsoumakas, I Katakis. *Multi-Label Classification: An Overview*.