

## Using Ontology for Goal-Based Query to Understand User Trends on Social Media: A Case of Twitter

<sup>1</sup>Tengku Adil Tengku Izhar, <sup>2</sup>Trieu Minh Nhut Le, <sup>3</sup>Ahmad A Alzahrani

<sup>1</sup>Universiti Teknologi MARA, UiTM Selangor, Malaysia

<sup>2</sup>Saigon University, Ho Chi Minh City, Vietnam

<sup>3</sup>King Abdulaziz University, Jeddah, Saudi Arabia

### Abstract

*Social media has a potential to be used as a professional and social networking to share interest. Social media such as Twitter can be used by users to share similar interests on the Internet. Despite the various existing methodologies on ontology, there are many opportunities that need to be explored. The aim of this paper focuses on structuring big data to set up a goal based on ontology to show the dependency relationship between social data and goal. At this stage of the research, we examine user trends on social media based on certain query as this query is defined as a goal. We use NodeXL to obtain and analyse data from Twitter. The results discuss the relationship for data from Twitter in relation to the specific goal. The outcome of this paper will contribute to a significant achievement that provides an important innovation in research methods in big data era to trace how data flows across the social media and how to analyse this data.*

### 1. Introduction

Today, there are tons of data collected from devices all over the world. Devices such as mobile phones, laptops, tablets and even video game such as Play Stations are connected to the Internet. These devices transmitted millions of data everywhere in the world and it highlights the development of big data. Billions of devices will be connected to the Internet and soon hundreds of billions of devices. As these devices connected to each other, these connected devices become an intelligent systems of systems. Intelligence systems of systems can share data and analyze it over cloud and they can transform our businesses, our lives and our world in countless ways. For example, improving the medical outcomes and building better product faster with lower development cost.

With the advance of mobile application, the way we access the Internet has also changed. The number of electronic devices that can be connected to network services has increased dramatically. The Internet connectivity is also possible beyond these devices. Every object around us connected, collecting, processing and sharing information. For

example, running shoes can upload the time, speed and distance of your run to your devices through Internet. As a result, it grows a memory capacity that has fueled the development of in-memory big data management and processing [1]. With the development of smart devices and cloud computing, more data can be collected from various sources and can be analyzed in an unprecedented way [2].

Nothing in technology stands still especially in the world of data and analytics. Systems built just a few years ago are starting to buckle under the explosion of data and the changing query needs of business users. Social media has now become an important medium of communication and interaction tools for social networks [3]. Social media is also important for business platform that can influence the corporative environment [4] [5] [6]. For example, social networks involve agents in creating and processing information for knowledge network [7]. At the same time, the role that causality can play in social network analysis is unclear [8]. Therefore, it is important to examine the flow of data in social media and to retrieve relevant data from large amount of it. In this paper, an ontology is apply to develop the relationship of data in social media. This relationship is important to capture data that can be considered relevant for certain query.

Ontology works as a type of relationship to represent the dependency relationship of data from the datasets and dependency relationship between data and goal. Despite the various existing methodologies on ontology[11] [12] [13] [14], there are many opportunities that need to be explored. The aim of this paper focuses on structuring the big data to set up a goal based on ontology to show the dependency relationship between social data and goal. The process consists of identifying which data are relevant to the goal. The paper examines user trends on social media based on certain query as this query is defined as a goal. The process is expected to increase precise relevant information from social media content. The research will make significant achievement that provides an important innovation in research methods in big data era to trace how data flows across the social media and how to analyse this data. Big data require increasingly more advanced

capabilities to find patterns in the inherent complexity for certain goal.

The ideas of using an ontology and visual structuring in organization applications were discussed in many works and now are implemented in many sectors [9] [10]. However, much of the research in this field did not receive much attention in the literature that incorporates big data for social media by setting up a query as a goal. An ontology provides explicit and formal specifications of knowledge, especially implicit or hidden knowledge [18]. By incorporating the big data, an ontology make the process to identify the relevance of data more easily consumable to address which data from the datasets are more important to evaluate certain goal.

The remainder of this paper is organized as follows. Section 2 discusses the background and challenges. Section 3 discusses the big data technology. Section 4 is the introduction of the ontology. Section 5 is the case study. A general discussion and future work are given in Section 6. The final section contains some concluding remarks.

## 2. Background

Government agencies and large, medium and small private enterprises in many domains, such as engineering, education, manufacturing, are drowning in an ever-increasing deluge of data. Companies like Google, eBay, LinkedIn and Facebook were built around big data from the beginning [19]. Big company like Apple's Corporation has the amount of data that is generated has risen steadily every year. For example, with the arrival of Apple's Watch, Apple presumed millions of people who will soon be using it for everything from monitoring their heart rate to arranging their social calendar to remote controlling their home entertainment [20]. This will doubtlessly bring with it a tsunami of data. Apple's Corporation itself has operational storage capacity from kilobytes to terabytes [21].

In today's competitive marketplace, executive leaders are racing to convert enterprise insights into meaningful results. Successful leaders are infusing analytics throughout their enterprises to drive smarter decisions, enable faster actions and optimize outcomes. For example, the IBM Institute of Business Value surveyed 900 business and IT executives from 70 countries. The results show that leaders are 166 percent more likely to make most decisions based on data [22]. Another good example is National Australia Bank (NAB) wanted to eliminate inconsistencies arising from storing data in 34 different financial and operational systems [23]. However, they lacked consistency in how to maintain the data. Therefore, it is very difficult for them to produced results in consistent manners that lead to their goal.

## 2.1. Challenges

In big data's science, people have access to more data in a single day rather than most people in the previous decade. Big data provides significant opportunities for enterprises to impact a wide range of business processes in organizations [24]. Organizations can create huge volume of data in their daily business activities. The problem though is that, this data was created and captured in many different formats, which make it almost impossible to understand the existing relationships between different data. As a result, this huge volume of data may get redundant and hard to identify which data is relevant to the goal. Big data may be created in petabytes or exabytes scale and much of which cannot be integrated easily. For example, the government agencies, the large, medium and small private enterprises in many domains such as engineering, education, manufacturing, are drowning in an ever-increasing deluge of data.

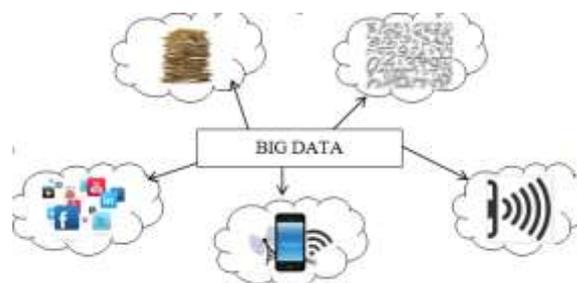


Figure 1. Creation of Big Data

Even though data scientists were trained to analyze data, but with the huge volume of data generated everyday makes it harder to identify which data is relevant. As a result, it poses an issue on how to effectively utilize this data to support decision-making processes [24]. Big data will be important to the organizations because more data can create more accurate analysis which can lead to more confident decision-making.

Firms and other organizations have been using large databases and analytics for the last couple of decades. Transactions are stored in data warehouses and analyzed with data-mining algorithms to extract insights [25]. Taking advantage of big data opportunities is a challenge for organizations [26]. In order to ensure the effectiveness of the data, organization needs to be able to store data reliably across a number of databases. Once data need to be distributed, organization needs a way to get it out again and they need to identify which data is needed, assemble it and analyse it. The challenge is how to capture this data to be considered relevant for the specific organization activities because determining relevant data is a key to delivering value from massive amounts of data as shown in Figure 2. The

real issue is not how the organization acquiring large amount of data but how they do with the data that counts [19]. Thus, having an ability to analyse the data in a timely fashion can ensure a competitive edge for efficient decision-making. However, the trustworthiness of data in relation to the goals is often questionable due to the huge amount of data especially from social media. The technologies and concepts behind big data can allow us to achieve a variety of objectives.

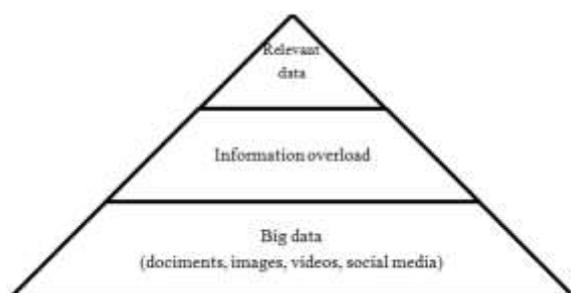


Figure 2. Relevant data from large amounts of data

### 3. Integration of big data technology

The implementation of big data highlights the development of big data analytics. Big data analytics could be used to examine large amounts of data of a variety of types to discover useful information. Such information can provide a competitive advantage to help better decisions. The primary goal of big data analytics is to help companies make better business decisions by enabling data analysts to analyze huge volumes of transaction data which remains an unresolved challenge for conventional business intelligence programs.

Big data analytics can be done with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics and data mining. However, the unstructured data sources used for big data analytics may not fit in traditional data warehouses. Furthermore, traditional data warehouses may not be able to handle the processing demands posed by big data. As a result, a new class of big data technology has emerged and is being used in many big data analytics environments. For example, the technologies associated with big data analytics include Hadoop, MapReduce and Hortonworks. These technologies form the core of an open source software framework that supports the processing of large data sets across clustered systems as shown in Figure 3.



Figure 3. Integration of big data technologies (adapted from Davenport [19])

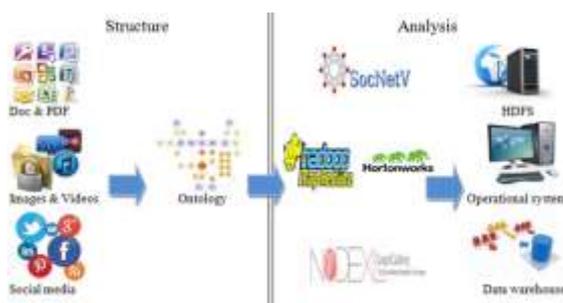


Figure 4. Integrating an ontology for big data

### 4. Research design

#### 4.1. A need for an ontology

In order to create a domain ontology, one can adopt various methods, like extending existing ontologies or developing the ontology from the ground up [27]. The contribution of an ontology is to improve the creation of model ultimately takes place through the defined goals [24].

Ontology architecture presented in Figure 5 consists of filtering and application. Filtering is a process to filter the data from the datasets by referring to the certain query. In this paper, we refer query as a goal. We develop the dependency relationship for this data in relation to the goal. For example, if we want to examine the number of tweet about PhD, then we will examine number of people who tweet or mention the word PhD in their profile. Application is a stage to use analytics to analyse the dependency relationship between data and goal. In this paper, an ontology is designed for big data to:

- Be applicable in a wide range of domain,
- Successfully structure the dependency relationship of data from the datasets,
- Successfully structure the dependency relationship between data that relate to the goal.

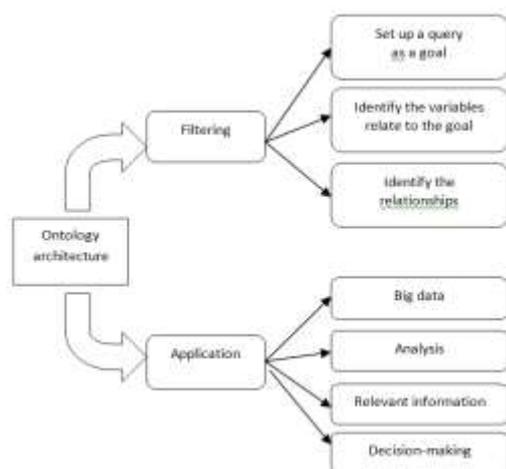


Figure 5. Ontology architecture for goal query

#### Filtering:

- Set up a query as goals

Set up the query for certain priorities. The query becomes our goals. For example, using Twitter, we want to identify if anyone in our network have mention or talk about certain issue.

- Identify the variables that relate to the goals

Identify the possible variables to assist the query. For example, Twitter has few variables such retweet, message, tweet, followers and following. These variables will assist users to evaluate the query they have on certain priorities.

- Identify the relationships

Based on the goals, we identify the relationships to be evaluated. Relationships such as people in the network who have replies, follows or mentions to the certain issues. It will be represented as different queries.

#### Application:

- Analytics process

Analyse the goals using the tools that is suitable to the data and goal. Tools such as Hadoop, Map Reduce and Hortonworks are popular technology for big data. Tools such as NodeXL and Social Networks Visualizer are popular to analyse big data in the context of social media.

- Relevant information and decision-making

The analysis process will retrieve relevant information to fulfill the information we needed for the queries. The information will drive the decision-making process.

## 4.2. Ontology for goals query

Despite many research efforts and established model for the evaluation of goal using an ontology [12, 13, 24], they have not yet been systematically applied to set up a query as a goal for social media. In this paper, several structures that were proposed in the previous models are combined, as shown in

Table 1. These models are adapted as a reference to set up goal query based on an ontology. However, the scope of the proposed ontology in this paper do not cover all the organizational processes as discussed in Sharma and Osei-Bryson [13], Fox et al. [12] and Rao et al. [10]. This is because we do not focus on the organizational structure but we extend the relationship between social data and goal.

Fox et al. [12] focused on structuring the linkage between organizational structure and behavior. This is critical for enterprise model development. However, the authors do not emphasize any organizational resources such as data and information but they focus on the roles and activities within the organization. Meanwhile, Sharma and Osei-Bryson [13] developed a framework for an organizational ontology in an effort to increase an understanding of the business. However, the authors do not specifically identify the relationship between organizational resources, such as data and the organizational goals. In this model, the authors adapted the work of Fox et al. [12], where the authors discussed the physical resources and role of the organizational model.

Recently, Rao et al. [10] developed an organizational ontology in order to build a knowledge map within the organization. The structure includes the flow of knowledge within the organization in the context of knowledge sharing and knowledge storage. In this model, the authors discussed the organizational resources, as in Sharma and Osei-Bryson [13]. Another aspect that is similar to Sharma and Osei-Bryson's work is that both models include business processes. However, Rao et al. [10] discussed business processes from the organizational goals point of view and Sharma and Osei-Bryson [13] discussed business processes from the organizational activity point of view. Most of these studies focused on the organizational structure and performance. In Izhar et al. [24], the authors developed the organizational goals ontology that consists of organizational goals, sub-goals, and organizational data. They developed the dependency relationship for the organizational goals and dependency relationship between organizational data and organizational goals. However, they evaluate the organizational goals by identifying the organizational goals first and then they identify the organizational data that relate to the organizational goals.

Table 1 shows the results from the previous models on the organizational goals using an ontology but none of these studies focus on structuring the big data in the context of social media. In this paper, we extend the organizational goals ontology from Izhar et al. [24] to set up goal for social media. Therefore, we can develop the relationship between social data and goal.

Table 1. Previous scope of the existing models on goal based on an ontology

Authors	Organizational goals ontology		Resources		Develop the relationships
	Organizational goals	Sub-goals	Organizational data	Social media	Big data
Fox et al. [12]					
Sharma and Ouel-Beyrou [13]	✓	✓	✓	✓	✓
Bao et al. [10]	✓	✓	✓	✓	✓
Labat et al. [14]	✓	✓	✓	✓	✓

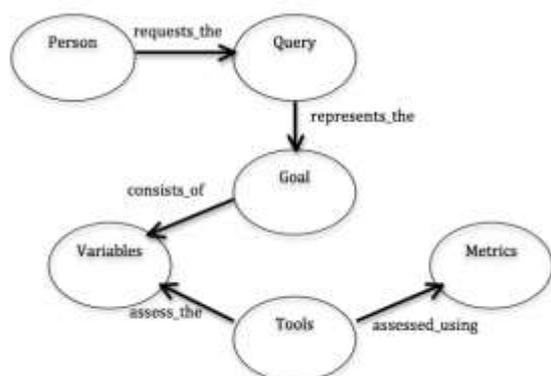


Figure 6. An ontology for goal query on social media using Visual Paradigm

Based on Table 1, we develop an ontology for goal query on social media, as shown in Figure 6. In Figure 6, we capture data from large database of social media by structuring this data to identify the dependency relationship of data that relate to the specific query. The relationships are proposed as follows;

- requests\_the (person, query); as person requests the query.
- represents\_the (query, goal); as query represents the goal.
- consists\_of (goal, variables); as goal consists of variables.
- assess\_the (tools, variables); as tools assess the variables.
- assessed\_using (tools, metrics); as tools assessed using metrics.

### 5. A case of Twitter

In this paper, we use Twitter as information source [28]. Twitter is a micro blogging site in which users can post updates (tweets) to friends (followers and following) [29]. We apply data from Twitter to test our ontology. However, Twitter users create high volume of data and make it difficult to analyse this data for certain query. For example, 250 million tweets per day create high amounts of data and the amounts are increasing every day. A reliable analysis approach for this data is a big challenge. Therefore, it

is important to structure this data by setting up a goal.

### 5.1. Ontology for Twitter

We select one username (@Write4Research) as an example. Write4Research is a Twitter profile that discuss on research activity and it is popular among researchers and PhD students. We set up a goal to examine the trend of people (followers and followings) who mention the word “PhD” in their tweets. In this example, there are three main variables in Twitter. There are other attributes that relate to these variables such as “favorites” and “message” but at this stage we only apply these variables to analyse the goal. The variables are:

- Username (@Write4Research)
- Followers
- Following
- Tweets

In this network, somewhere there might be someone who mentions the word “PhD” based on three relationships. The relationships are:

- “tweet\_the” relationship.
- “replies\_to” relationship.
- “mentions\_the” relationship.

The relationships shows when the person in this network who tweet, mentions or replies to one another tweets about PhD. We analyse these relationships that represents three different queries. We develop an ontology for these queries as follows:

Query 1: People who tweet about PhD

- tweet\_the (Followers, PhD), as followers tweet PhD.
- tweet\_the (Following, PhD), as following tweet PhD.

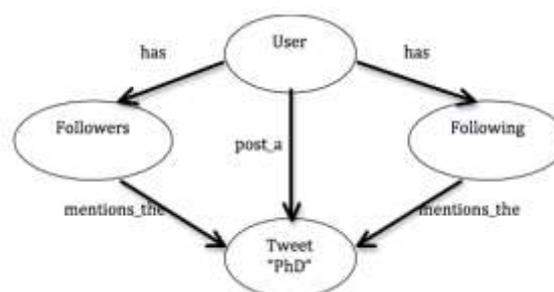


Figure 7. Ontology for query 1 for people who follows the tweet about PhD

Query 2: People who replies to the tweets about PhD.

- replies\_to (Followers, PhD), as followers replies to PhD.

- replies\_to (Following, PhD), as following replies to PhD.

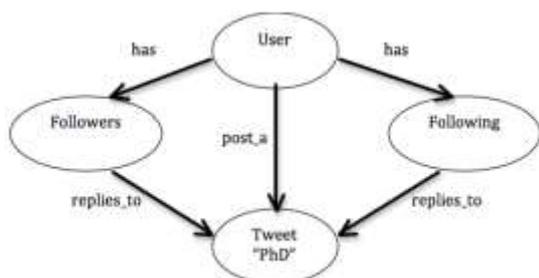


Figure 8. Ontology for query 2 for people who replies to the tweet about PhD

Query 3: People who mention the tweets about PhD.

mentions\_the (Followers, PhD) as followers mentions the PhD.

mentions\_the (Following, PhD) as following mentions the PhD.

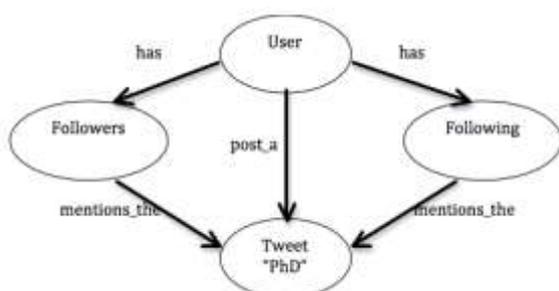


Figure 9. Ontology for query 3 for people who mentions the tweet about PhD

## 5.2. Sampling and data collection

To identify participants, we used the sampling pool of Twitter that appears in the NodeXL, a software tool that import data from outside data providers. It is an extendible toolkit for network overview, discovery and exploration implemented as an add-in to the Microsoft Excel 2007 spreadsheet software. NodeXL is applied to retrieve data from social media and import this data. NodeXL demonstrate data analysis with a social media data sample drawn from an enterprise intranet social network. We apply NodeXL to visualize the ontology for three different queries. When it comes to collecting, computing, analyzing, and acting on social data, technical challenges are quite different because number of social data always increase and make it difficult to evaluate. For example, today data might be important but tomorrow this data might not be important anymore. Therefore, we filter this data based on certain date and for each relationship.

**5.2.1. Filter** This section will provide steps to demonstrate how we filter data from large amount of data from the social media to allow us to evaluate

specific query in relation to the goals. These preferences are used to configure the steps in filtering the data from Twitter using NodeXL.

1. Import from Twitter users network, as shown in Figure 10.

1.1. It optionally clear the NodeXL workbook, then get the network of specified Twitter users.

2. Specify the Twitter users with specific username.

2.1. We interested in username @Write4Research.

3. Import basic network plus followers and following who replies, mentions and tweet.

3.1. Limit it to 100 recent tweets per user.

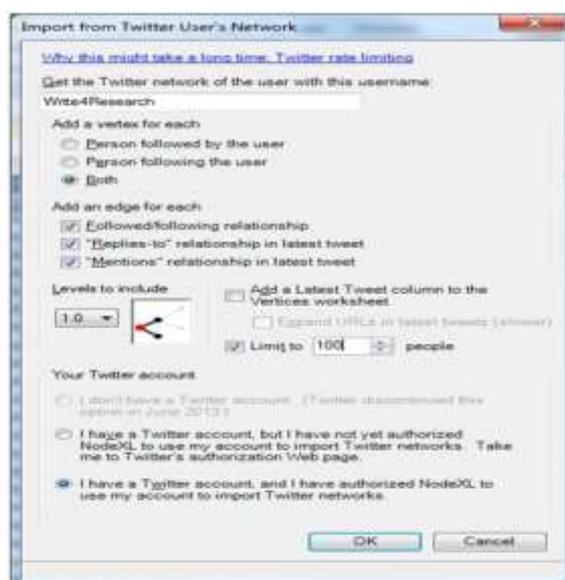


Figure 10. Import tweet username

4. Import from Twitter search network, as shown in Figure 11.

4.1. It optionally clear the NodeXL workbook, then get the network of people who tweets certain specified word.

5. Search for the tweets that match to the specific query.

5.1. We search for PhD.

6. Import basic network to specifically show who replied or mentioned in the tweets.

6.1. Limit to 100 tweets.

7. Filter by relationships (tweet, mentions and replies).

The step-by-step guideline explains how we select data from huge volume of social data that relate to the query. Using NodeXL as a tool to

capture this data, this guideline provide systematic steps for domain experts to capture which data they

other case, domain experts might use different tool to capture this data and the process to identify relevant data might be different.

want to evaluate to assist their decision-making in relation to the organizational goals. However, in



Figure 11. Import tweet from username network

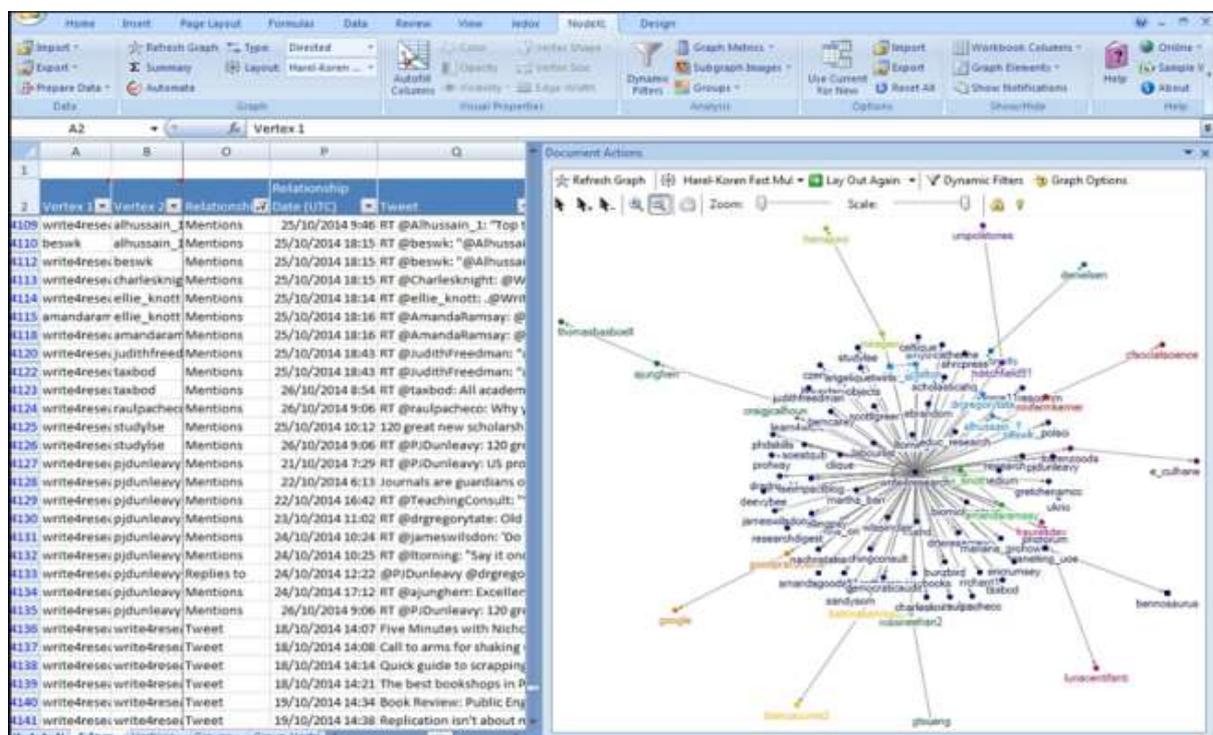


Figure 12. Tweet that match to the queries in Write4Research network

**5.3.1. Analysis and results** In this case study, we examine user trends on Twitter by visualising the relationships in Write4Research’s network. The relationship is evaluated based on the vertex in-degree and out-degree graphs. The overall graph metrics is shown in Appendix A.

**5.3.2. Vertex in-degree (directed graphs only)** In a directed graph, a vertex's in-degree is the number of incoming edges incident to the vertex. In an undirected graph, in-degree is undefined and is not calculated. A self-loop in a directed graph is counted once as an incoming edge (in-degree) and once as an outgoing edge (out-degree).

**5.3.3. Vertex out-degree (directed graphs only)** In a directed graph, a vertex's out-degree is the number of outgoing edges incident to the vertex. In an undirected graph, out-degree is undefined and is not calculated. A self-loop in a directed graph is counted once as an outgoing edge (out-degree) and once as an incoming edge (in-degree).

The Figure 12 shows the entire relationships in Write4Research network. The graph shows the complexity of this relationship in order to retrieve relevant information for certain query.

The issue is to identify different relationship for people who tweets, replies and mentions to each other. In this figure, the detail of this data is imported to the spreadsheet to be analysed. However, we only analyse users trend who tweet about PhD by visualizing the graph for different relationships. The results in Table 3 show the overall outcome for query 1.

The Figure 13 shows the result for query 1, that shows the relationship of people who start the tweet about PhD. Figure 13 shows that they are not many people who tweet to PhD topic on the selected date. The results in Table 3 show the overall outcome for query 1.

The Figure 14 shows the result for query 2, that shows the relationship of people who retweet about PhD. The graph shows the relationship that linked to different profile. Figure 14 shows the number of people who retweet to PhD topic on the selected date. The result in Table 4 shows the overall outcome for query 2.

The Figure 15 shows the result for query 3, that shows the relationship of people who replies to the tweet. The graph also shows the relationship that linked to different profile. They are not many people who reply to the tweet on the selected date. The result in Table 5 shows the overall outcome for query 3.



Figure 13. Query 1 for people who tweet about PhD

Table 2. Summary of tweet that match to the queries in Write4Research network

Graph Metric	Value
Graph Type	Directed
Vertices	102
Unique Edges	80
Edges With Duplicates	59
Total Edges	139
Self-Loops	13
Reciprocated Vertex Pair Ratio	0.0126582278481013
Reciprocated Edge Ratio	0.025
Connected Components	32
Single-Vertex Connected	11

Components	
Maximum Vertices in a Connected Component	25
Maximum Edges in a Connected Component	73
Maximum Geodesic Distance (Diameter)	3
Average Geodesic Distance	1.833873
Graph Density	0.007821782
Modularity	0.562393
NodeXL Version	1.0.1.333

Table 3. Overall metrics of query 1 for people who tweet about PhD

Graph Metric	Value
Graph Type	Directed
Vertices	12
Unique Edges	11
Edges With Duplicates	2
Total Edges	13
Self-Loops	13
Reciprocated Vertex Pair Ratio	Not Applicable
Reciprocated Edge Ratio	Not Applicable
Connected Components	12
Single-Vertex Connected Components	12
Maximum Vertices in a Connected Component	1
Maximum Edges in a Connected Component	2
Maximum Geodesic Distance (Diameter)	0
Average Geodesic Distance	0
Graph Density	0
Modularity	0.281065
NodeXL Version	1.0.1.333

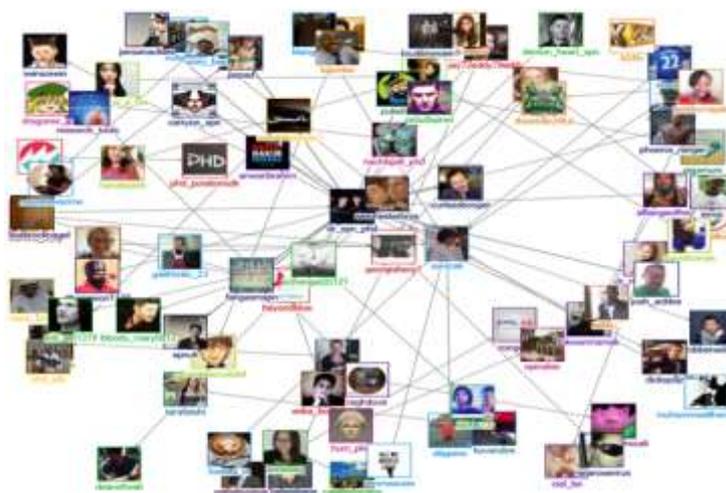


Figure 14. Query 2 for people who mentions (retweet) about PhD

Table 4. Overall metrics of query 2 for people who mentions the tweet about PhD

Graph Metric	Value
Graph Type	Directed
Vertices	85
Unique Edges	61
Edges With Duplicates	53
Total Edges	114
Self-Loops	0
Reciprocated Vertex Pair Ratio	0
Reciprocated Edge Ratio	0
Connected Components	20
Single-Vertex Connected Components	0
Maximum Vertices in a Connected Component	25
Maximum Edges in a Connected Component	69
Maximum Geodesic Distance (Diameter)	3
Average Geodesic Distance	1.881564
Graph Density	0.009803922
Modularity	0.557749
NodeXL Version	1.0.1.333



Figure 15. Query 3 for people who reply to the tweet about PhD

Table 5. Overall metrics of query 3 for people who replies to the tweet about PhD

Graph Metric	Value
Graph Type	Directed
Vertices	19
Unique Edges	10
Edges With Duplicates	2
Total Edges	12
Self-Loops	0
Reciprocated Vertex Pair Ratio	0.1
Reciprocated Edge Ratio	0.181818182
Connected Components	9
Single-Vertex Connected Components	0
Maximum Vertices in a Connected Component	3
Maximum Edges in a Connected Component	3
Maximum Geodesic Distance (Diameter)	2
Average Geodesic Distance	0.585366
Graph Density	0.032163743
Modularity	0.666667
NodeXL Version	1.0.1.333

The graphs show interesting results for these three relationships. The results show that people in Write4Research's network prefer to retweet about PhD, as shown in Figure 14 and Table 4. It shows that they are interested to share about this topic in their network. We come out with this decision because majority of Write4Research followers are researchers.

## 6. Discussion and future work

A unique contribution of this paper is its perspective on how to develop the relationship between social data and goal. The relationship help the process to identify the goal, identify the relevant data that relate to the goal and analyse this relationship to retrieve relevant information to support decision-making.

### 6.1. Contribution

- Flexible to identify the goal: In this paper, we explained how to we set up the query as goal. The usage of ontology assists the flexibility to define the goal. By using this ontology, the process to identify the set of this goal becomes flexible.
- Flexible to identify the dependency relationship: Ontology developed the dependency relationship for social data based on query. At the same time, we developed the relationship between social data and goal. We proved this flexibility in the case study in which we developed the dependency relationship for data in Twitter. This flexibility assists the process to identify which data to be consider relevant to the query.
- Flexible to analyse the data: The main objective of data analysis is to evaluate data from the vast amount of datasets. In this paper, we used visual analysis to show the relationship for this data. The value from the graph allows us to examine the relationship in relation to the query. The results in the case study to show the dependency relationship of people who tweet, replies and mentions the word PhD.

### 6.2. Future work

Evidence has shown that ontology can be effectively developed these relationships. For example, despite the challenges in capturing relevant data from social media, filtering this data would be a better solution to store and analyse this data. As large volumes of social data are unstructured, data becomes more complex to analyse. NodeXL has been found to deal better with complex and unstructured social data, in which it can import social data in basic spreadsheet. However, this paper limited to social data only. Therefore, future work can be conducted to analyse this data. At the same

time, we can examine this data to see if data from social media can support organizational data for better decision-making. We extend the ontology to incorporate social media and organizational data, as shown in Figure 16.

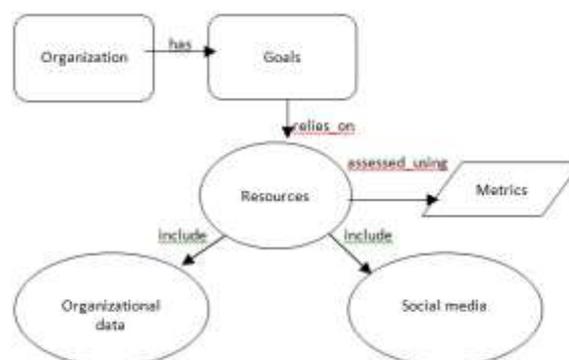


Figure 16. Future work

## 7. Conclusion

The Social Media such as Facebook and Twitter have become an indispensable part of our lives. In this paper, we have described the main features of ontology when developed the relationship for social data in relation to the goal. In addition, we have proposed an alternative approach to capture social data using NodeXL. We extend the application of the ontology for social media in relation to the goal by developing the relationship for this social data. This paper enables researchers to classify and evaluate existing research, to design scientific research, to identify the gaps and weaknesses prior to future research directions.

## 8. References

- [1] H. Zhang, G. Chen, B. C. Ooi, K.-L. Tan, and M. Zhang, "In-memory big data management and processing: A survey," *IEEE Transaction on Knowledge and Data Engineering*, vol. 27, pp. 1920-1948, 2015.
- [2] T. Huang, L. Lan, X. Fang, P. An, J. Min, and F. Wang, "Promises and challenges of big data computing in health sciences," *Big Data Research*, vol. 2, pp. 2-11, 2015.
- [3] N. A. Ghani and S. S. M. Kamal, "A Sentiment-based filtration and data analysis framework for social media," in *5th International Conference on Computing and Informatics ICOCI*, ed. Istanbul, Turkey, 2015, pp. 632-637.
- [4] L. Holtzblatt, J. L. Drury, D. Weiss, L. E. Damianos, and D. Cuomo, "Evaluating the uses and benefits of an enterprise social media platform," *Journal of Social Media for Organizations*, vol. 1, pp. 1-21, 2013.
- [5] L. E. Domianos, D. Cuomo, J. Griffith, D. M. Hirst, and J. Smallwood, "Exploring the adoption, utility and social influences of social bookmarking in a corporate

environment," presented at the International Conference on System Sciences, Hawaii, 2007.

[6] L. E. Damianos, D. L. Cuomo, and S. Drozdetski, "Handshake: A case study for exploring business networking for enterprise, inside and out," in International Conference on Human Computer Interaction, 2011.

[7] C. Roth and J. P. Cointet, "Social and semantic coevolution in knowledge networks," *Social Networks*, vol. 32, pp. 16-29, 2010.

[8] P. Doreian, "Causality in social network analysis," *Sociological Methods & Research*, vol. 30, pp. 81-114, 2001.

[9] G. Mansingh, K.-M. Osei-Bryson, and H. Reichgelt, "Building ontology-based knowledge maps to assist knowledge process outsourcing decisions," *Knowledge Management Research and Practice*, vol. 7, pp. 37-51, 2009.

[10] L. Rao, G. Mansingh, and K.-M. Osei-Bryson, "Building ontology based knowledge maps to assist business process re-engineering," *Decision Support Systems*, vol. 52, pp. 577-589, 2012.

[11] L. Rao, H. Reichgelt, and K.-M. Osei-Bryson, "An approach for ontology development and assessment using a quality framework," *Knowledge Management Research and Practice*, vol. 7, pp. 260-276, 2009.

[12] M. S. Fox, M. Barbuceanu, M. Gruninger, and J. Lin, "An organization ontology for enterprise modelling," in *Simulation organizations: Computational models of institutions and groups* AAAI/MIT Press, ed, 1998, pp. 131-152.

[13] S. Sharma and K.-M. Osei-Bryson, "Organization-ontology based framework for implementing the business understanding phase of data mining projects," in *International Conference on System Sciences*, Hawaii, 2008, p. 27.

[14] M. S. Fox, M. Barbuceanu, and M. Gruninger, "An organisation ontology for enterprise modeling: Preliminary concepts for linking structure and behaviour," *Computers in Industry*, vol. 29, pp. 123-134, 1996.

[15] J. Valaski, A. Malucelli, and S. Reinehr, "Ontologies application in organizational learning: A literature review," *Expert Systems with Applications*, vol. 39, pp. 7555-7561, 2012.

[16] M. B. Almeida and R. R. Barbosa, "Ontologies in knowledge management support: a case study," *Journal of the American Society for Information Science and Technology*, vol. 60, pp. 2032-2047, 2009.

[17] M.-C. Valiente, E. Garcia-Barriocanal, and M.-A. Sicilia, "Applying an ontology approach to IT service management for business-IT integration," *Knowledge-Based Systems*, vol. 28, pp. 76-87, 2012.

[18] J. Cho, S. Han, and H. Kim, "Meta-ontology for automated information integration of parts libraries," *Computer-Aided Design*, vol. 38, pp. 713-725, 2006.

[19] T. H. Davenport and J. Duche, "Big data in big companies," *International Institute for Analytics* 2013.

[20] B. Marr. Big data-as-a-a-service is next big thing. [Internet] Available:<http://www.forbes.com/sites/bernardmarr/2015/04/27/big-data-as-a-service-is-next-big-thing/2/> (Access date: 30 April 2015)

[21] J. Aluya. A case study of Apple Corporation's use of large datasets. [Internet] Available:[http://www.academia.edu/8112361/A\\_case\\_study\\_of\\_APPLE\\_Corporations\\_use\\_of\\_large\\_datasets](http://www.academia.edu/8112361/A_case_study_of_APPLE_Corporations_use_of_large_datasets) ( Access date: 30 April 2015)

[22] Y. Shani, "Addressing data storage challenges with a new family of solution," in *IBM Big data & Analytics Hub*, ed. IBM, 2015.

[23] D. Baum. National Australia Bank improve data quality, streamlines regulatory. [Internet] Available: <http://www.oracle.com/us/c-central/cio-solutions/information-matters/importance-of-data/information-management-1888241.html> (Access date: 30 April 2015)

[24] T. A. T. Izhar, T. Torabi, M. I. Bhatti, and F. Liu, "Recent developments in the organization goals conformance using ontology," *Expert Systems with Applications*, vol. 40, pp. 4252-4267, 2013.

[25] J. R. Galbraith, "Organization design challengers resulting from big data," *Journal of Organization Design*, vol. 3, pp. 2-13, 2014.

[26] M. Berber, E. Graupner, and A. Maedche, "The information panopticon in the big data era," *Journal of Organization Design*, vol. 3, pp. 14-19, 2014.

[27] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, "Ontology-based sentiment analysis of twitter posts," *Expert Systems with Applications*, vol. 40, pp. 4065-4074, 2013.

[28] A. S. Veenstra, N. Iyer, M. D. Hossain, and J. Park, "Time, place, technology: Twitter as an information source in the Wisconsin labor protests," *Computer in Human Behavior*, vol. 31, pp. 65-72, 2014.

[29] N. F. F. d. Silva, E. R. Hruschka, and E. R. H. Jr, "Tweet sentiment analysis with classifier ensembles," *Decision Support Systems*, vol. 66, pp. 170-179, 2014.