International Journal of Academic Research in Business and Social Sciences
2016, Vol. 6, No. 9
ISSN: 2222-6990

Retrieving Information from Social Media using Ontology

Tengku Adil Tengku Izhar, Mohd Shamsul Mohd Shoid
Faculty of Information Management, Universiti Teknologi MARA,
UiTM Selangor, Malaysia

Abdul Ismail Mohd Jawi
Faculty of Information Management
Universiti Teknologi MARA,
UiTM Sarawak,
Malaysia

Trieu Minh Nhut Le
Department of Computer Science and Computer Engineering,
Saigon University, Ho Chi Minh City, Vietnam

DOI: 10.6007/IJARBSS/v6-i9/2267 URL: http://dx.doi.org/10.6007/IJARBSS/v6-i9/2267

Abstract
People have access to more data in single day than most people that have access to data in the previous decade. This data is created in many forms and it highlights the development of Big Data. Big Data in organizations have transformed the way organizations across industries implement new approach to handle huge amount of data. Organizations rely to this data to achieve specific business priorities. The challenge is how to retrieve this data to be considered relevant for the specific organization activities because determining relevant data is a key to deliver information from massive amounts of data. The aim of this paper is to integrate organizational data and social data using an ontology to retrieve relevant information for efficient decision-making. We investigate how external data such as social media can support internal data such as organizational data in relation to the organizational goals. The results from the case study demonstrate how we incorporate social data and organizational data. This paper demonstrates that ontology provide a platform to integrate social data and organizational data.

Keywords: Big Data; Ontology; Organizational data; Social Media; Twitter

1. Introduction Big Data provides significant opportunities for enterprises to impact a wide range of business processes in organizations (Berber et al., 2014; Izhar et al., 2013; Kang & Sim, 2011; Sá et al., 2015). Although there are many studies conducted on Big Data in the context of the organizations (Berber et al., 2014; Galbraith, 2014; Grossman & Siegel, 2014; Hazen et al., 2014), there is still little debate these days on the role and importance of Big Data for efficient decision-making (Davenport & Dyche, 2013; Grossman & Siegel, 2014; Manyika et al., 2011).Even though there are many recent studies conducted on Big Data in the context of the organizations (Berber et al., 2014; Galbraith, 2014; Grossman & Siegel, 2014; Hazen et al., 2014), there is still little debate these days on the role and importance of Big Data for better decision making to support the goals of an organization (Davenport & Dyche, 2013; Grossman & Siegel, 2014). There is yet no consensus about how best to incorporate Big Data in the organizations and how the process of incorporating Big Data can identify the relevance of data to assist in decision-making process. The emergence of Big Data has led to a profound transformation on how organizations store, manage and analyse their data. Increased access to large-scale data enables an organization on how to capture relevant data, which can assist in their decision making process. It is now possible to manage the volume, velocity and variety of data using the tools and software, so the focus in this research is to shift towards how we can use analyses to find extra value from this data analysis in relation to certain organizational goals. This will make significant achievement that provides an important innovation in research methods in Big Data era to trace how data flows across organizations. At the same time, it will educate organizations on how to analyze data to support their decision-making process.

1.1 Purpose

The aim of this research is to develop a framework that underpins a seamless integration of organizational data and heterogeneous external data pertinent to the organization's focus area. In this era of Big Data science, it is critical for organizations and businesses to be able to embrace this new facility and to accurately integrate the knowledge-bases from multiple sources into the organizational information repository. The availability of a mechanism that allows seamless consolidation of knowledge from external sources will enrich the capability of the organization to make accurate decision-making. These heterogeneous external sources are growing very significantly in the last few years, especially due to the availability of wireless and mobile technologies, crowd-sourcing facilities, Internet of Things and sensor networks, as well as social media and web data. All these technologies generate huge amount of data and together they can be extracted to generate values to the organization and to establish situational awareness of the community or market trends.

The main complexity of establishing the above mentioned framework lies on the heterogeneity of the data and information sources. In addition, they all have different nature in terms of volume and velocity (e.g. sensor data and social media data may be large amount of streaming data, other data sets may be less dynamic), spatial relativity aspect (e.g. mobile data may have inherent spatial/GPS knowledge that can be extracted), and quality aspects (e.g. crowd-source data may be less reliable), etc.

The remainder of this paper is organized as follows. Section 2 is literature review. Section 3 discusses on the proposed framework. Section 4 is case study. Section 5 is metrics and data collection. Section 6 is discussion. The final sections contain the future works and concluding remarks

2. Literature review

Society is becoming increasingly more instrumented and as a result, organizations are producing and storing vast amounts of data. Managing and gaining insights from the produced data is a challenge and key to competitive advantage (Assuncao et al., 2015). Big data is a new way of thinking about enterprise data and how it can drive business value. The amount of data that is available to businesses is increasing, with social media and machine-to-machine as just two of the leading sources. Storage is getting cheaper and processing power is getting faster. The central role of business services in today's enterprises, and the more complex architecture through which they are delivered, make it essential to manage big data solutions from a business perspective. Business perspective focuses on business objectives and benefit, and prioritizes resources and activities according to the needs of the business. In this way, structuring the big data can ensure optimal relevance of data for more effective decision-making (Izhar et al., 2013).

With the development of smart devices and cloud computing, more and more data can be collected from various sources and can be analyzed in an unprecedented way. The huge social and academic impact of such developments caused a worldwide buzz for big data (Huang et al., 2015). For example, social media have been adopted by many businesses. More and more companies are using social media tools such as Facebook and Twitter to provide various services and interact with customers. As a result, a large amount of user-generated content is freely available on social media sites (He et al., 2013).

Social media has been a popular topic among scholars spanning several disciplines including communication, psychology, sociology and business. The bulk of existing academic literature on social media has been published in just the last few years and has focused on the social processes of social media and its effects in areas such as marketing, politics, health communication, and education (McIntyre, 2014). Social media platform such as Twitter has stormed onto the social media scene not only as an individual communication device but also as an information dissemination platform. People on social media express opinions on different topics (Silva et al., 2014).

Currently, the majority of social media studies focus on individual companies or organizations. There are few studies performing social media competitive analysis on the leading companies in an industry in a systemic way (He et al., 2013). Social media can help decision makers to ensure efficient solutions to the problems raised (Silva et al., 2014). However, the trustworthiness of this social data is often questionable due to the huge amount of data created in social media.

In recent years, the valuable knowledge that can be retrieved from petabyte scale datasets that can led to the development of solutions to process information based on parallel and distributed computing (Polato et al., 2014). This paper aims to develop a framework to extract value from social data based on mapping and filtering. In order to achieve this framework, we adopt data processing approach based on MapReduce. The data processing strategy employed by MapReduce consists of two primitive functions: Map and Reduce. Behind this simple abstraction is a single fixed data flow. A MapReduce job is divided into Map and Reduce tasks, and assigned to idle slots of workers according to these two stages (Polato et al., 2014).

In recent years, the rapid development of Internet, Internet of Things, and Cloud Computing have led to the explosive growth of data in almost every industry and business area. Big data has rapidly developed into a hot topic that attracts extensive attention from academia, industry, and governments around the world. There are many challenges in harnessing the potential of big data today, ranging from the design of processing systems at the lower layer to analysis means at the higher layer, as well as a series of open problems in scientific research. Big data processing systems suitable for handling a diversity of data types and applications are the key to supporting scientific re-search of big data (Jin et al., 2015).

Social media are transforming the way information travels within and between networks of individuals (Gangadharbatla et al., 2014). Although the research on social networks dates back to early 1920s, nevertheless, social media analytics is a nascent field that has emerged after the advent of Web 2.0 in the early 2000s. The key characteristic of the modern social media analytics is its data- centric nature. Social media analytics refer to the analysis of structured and unstructured data from social media channels. Social media is a broad term encompassing a variety of online platforms that allow users to create and exchange content. User-generated content (e.g., sentiments, images, videos, and bookmarks) and the relationships and interactions between the network entities (e.g., people, organizations, and products) are the two sources of information in social media (Gandomi & Haider, 2015).

Social media have profoundly changed our lives and how we interact with one another and the world around us (Qualman, 2009; Safko & Brake, 2009). Recent research indicates that more and more people are using social media applications such as Facebook and Twitters for various reasons such as making new friends, socializing with old friends, receiving information, and entertaining themselves (Kaplan & Haelein, 2010; Keckley, 2010; Park, Kee, & Valenzuela, 2009; Raacke & Bonds-Raacke, 2008). Social media analysis will extract value from vast amount of social media data to detect and discover new knowledge to understand how industry is changing, and use the findings and improved understanding to achieve competitive advantage against their competitors (Governatori & Iannella, 2011; He et al., 2013). Social media competitive analysis allows a business to gain possible business advantage by analyzing the publicly available social media data of a business and its competitors (He et al., 2013). As social media have become a topic of interest for many industries, it is important to understand how social media data can be harvested for decision-making (He et al., 2013).

With the development of smart devices and cloud computing, more and more public data can be collected from various sources and can be analyzed in an unprecedented way. The huge social and academic impact of such developments caused a worldwide buzz for big data. Data flow is an ordered sequence which is consecutive, high-speed, infinite and time varying. It's also of great importance in internet management, internet security and internet experiment. However, with the rapid development of internet technology, the number of internet applications and users keeps rising, and the internet data is growing exponentially (Zhi et al., 2011). As a result, there is a stricter requirement about the efficiency, expandability and stability of the data flow in social media.

3. Framework

The framework aims to resolve the issue in identifying and evaluating relevant data for better decision-making that covers the characteristics of good quality relevant data.

3.1 The role of ontology

The ideas of using an ontology and visual structuring in organization applications were discussed in many works and now are implemented in many sectors (Almeida & Barbosa, 2009; Mansingh et al., 2009; Rao, Mansingh, & Osei-Bryson, 2012; Valaski et al., 2012; Valiente et al., 2012). However, much of the research in this field did not receive much attention in the literature on incorporating the Big Data for social media to assist the organizations with the decision-making process in relation to the organizational goals.

An ontology provides explicit and formal specifications of knowledge, especially implicit or hidden knowledge (Cho et al., 2006). 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 in evaluating the goals. The outcome of this paper can establish an analytics of Big Data structure for the organizations to ensure that analytics processes are supported by the specific organizational priorities. The contribution of an ontology is to improve the creation of model ultimately takes place through the organizational goals and it works as a type of relationship to represent the dependency relationship between data and organizational goals, as shown in Figure 1.

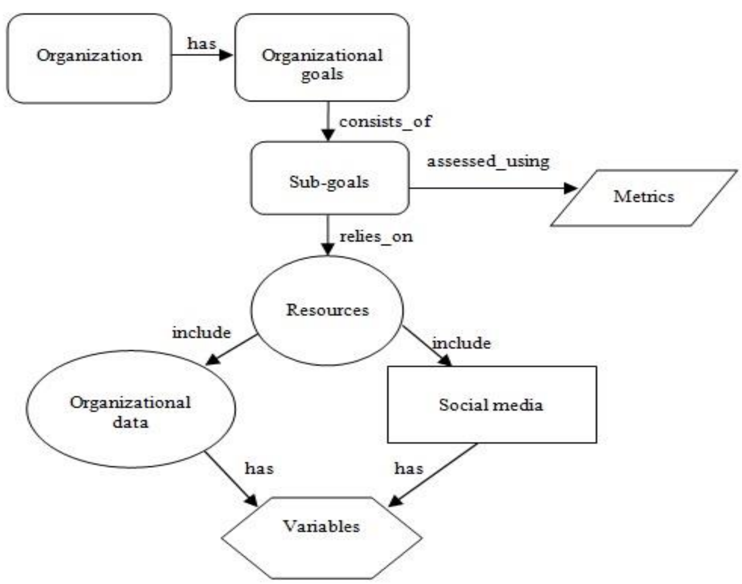


Figure1. Extended version of the organizational goals ontology (adapted from Izhar et al. (2013))

Despite the various existing methodologies to evaluate the organizational process based on an ontology (Fox et al., 1996; Fox et al., 1998; Mansingh et al., 2009; Rao et al., 2012; Rao et al., 2009; Sharma & Osei-Bryson, 2008), this paper focuses on structuring the relationship between social data and organizational data in relation to the organizational goals. This process consists of identifying which data are relevant in achieving the organizational goals that will be used by domain experts and entrepreneurs who contribute to the decision-making process. They are also responsible for identifying to what extent the organizational goals have been achieved.

4. Case study

This case study is presents to evaluate the relationship between organizational data and social data. The aim is to investigate how data from social media can incorporate with the organizational data for better decision-making in relation to the organizational goals. In order to achieve this aim, we present a case study from La Trobe University Student Support Services. We evaluate internal data, which is data that have been collected in the university and external data, which is data that have been collected from social media. We apply data from Twitter in relation to the case study goal.

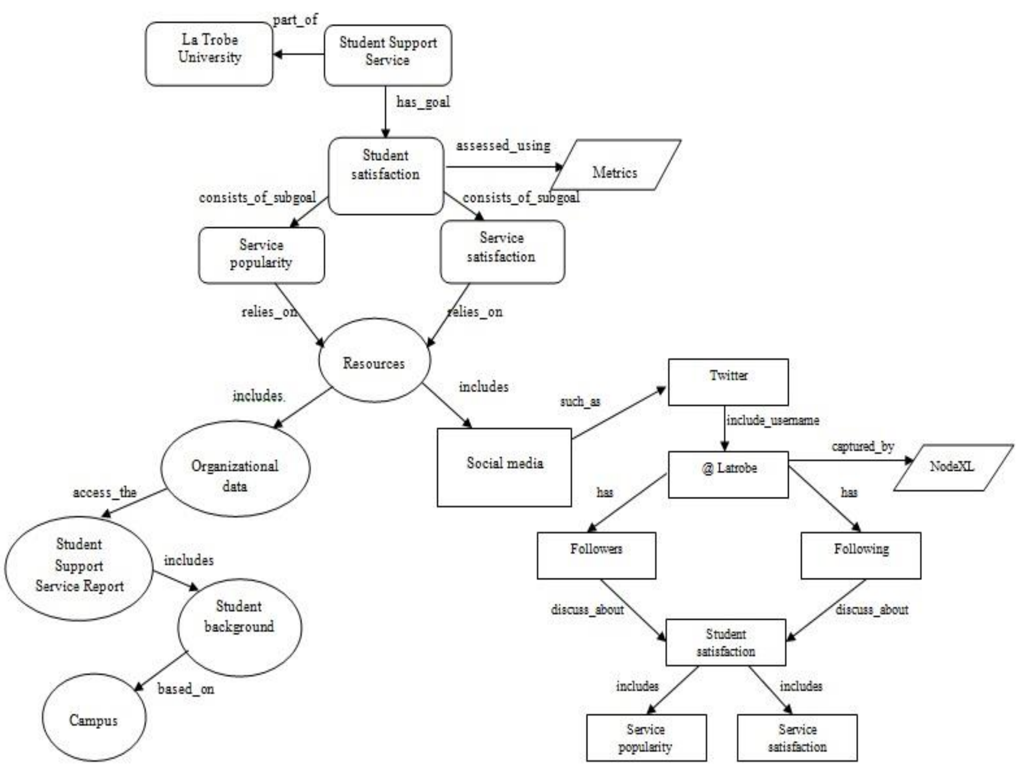


Figure 2. Ontology for La Trobe University.

4.1 Dependency relationships for the goals:

We identify the relationships for the goal and sub-goals. In Figure 2, Student Support Service is part of La Trobe University. Student Support Service has a goal. The goal is to evaluate the level of student satisfaction. This goal is assessed using metrics. Student satisfaction consists of sub-goals. The sub-goals are service popularity and service satisfaction. The sub-goals rely on resources. Resources are organizational data and social data. We identify the relationships are:

- *part_of* (student support service, La Trobe University)
- *has_goal* (student support service, student satisfaction)
- *assessed_using* (student satisfaction, metrics)
- *consists_of_sub-goal* (student satisfaction, service popularity)
- *consists_of_sub-goal* (student satisfaction, service satisfaction)
- *relies_on* (service popularity, resources)
- *relies_on* (service satisfaction, resources)

4.2 Dependency relationships for organizational data:

We identify the relationship for organizational data. Organizational data access the Student Support Service Report and this report includes student background. In this case study, we only evaluate the background based on campus. The relationships are defined as:

- *access_the* (organizational data, student support service report)
- *includes* (student support service report, student background)
- *based_on* (student background, campus)

4.3 Dependency relationship for social data:

In order to support the evaluation of organizational data in relation to the goal in this case study, we evaluate data from social media. In this case study, we present data from Twitter. Data are selected based on LaTrobe username. People in LaTrobe network discusses about student satisfaction. In this user network, someone might mention or tweet the word "student satisfaction". Then, we filter our search from student satisfaction to service popularity and service satisfaction. This includes people who tweet, mentions and replies the word student satisfaction.

The relationships shows when the person in the user network who tweet, mentions or replies to one another tweet about student satisfaction. In this case study, we summarize all these relationships as discussed about. It means people tweet, replies and mention about certain topic. The relationships are defined as:

- *such_as* (social media, Twitter)
- *include_username* (Twitter, LaTrobe)
- *has* (LaTrobe, followers)

- *has* (LaTrobe,following)
- *discuss_about* (followers, student satisfaction)
- *discuss_about* (following, student satisfaction)
- *includes* (student satisfaction, service popularity)
- *includes* (student satisfaction, service satisfaction)

5. Metrics and Data Collection

5.1 Metrics

In this section we define the metrics. This section identifies the different weights which were assigned to the dependency variables to measure the level of student satisfaction in the La Trobe Student Support Services based on service satisfaction and service popularity. The case study used in this experiment provides an approach to show how the data relating to the case study goal can be analysed. However, we are mindful of the fact that domain experts and entrepreneurs might want to analyse data in a different way to the way we have undertaken the analysis in this case study, which would require a different approach to define the metrics.

The overall student satisfaction based on service satisfaction and service popularity is calculated based on total rank and percentage using the following metrics:

$$\text{Total rank} = (\text{Total} \times \text{Rank})$$

$$\text{Percentage} = \left(\frac{\text{Total}}{\text{Total rank}} \times 100\right)$$

5.2 Evaluate of Organizational Data:

We applied dataset from La Trobe University Student Support Services Experiment Report in 2011. A survey was conducted online and the students were invited to participate by completing an anonymous questionnaire, as guaranteed confidentiality helps ensure that the true concerns of the students are identified. The survey firstly asked students to provide some demographic information regarding the La Trobe University Services which were considered critical to the satisfaction of students. Students were asked to indicate whether:

- they had not used the services but were aware of them.
- they had used the services but believed the services could be improved.

Table 1.Total Score Of Service Population And Service Satisfaction Based On Campus

Campus	Service popularity	Rank	Service satisfaction	Rank
Bundoora	68	2	32	5
Bendigo	67	3	33	4
Albury Wodonga	70	1	30	6
Mildura	66	4	34	3
Shepparton	63	5	37	2
City	59	6	41	1
Total/ Total rank	393	1340	207	689
Percentage	29		30	

In this dataset, we selected 11 different services provided by La Trobe University. It is important to note that we do not mean to imply that the entire services are not important but we suggest that these numbers of services are enough to test the flexibility of the framework with respect to the La Trobe University Student Support Services. A value is assigned to each La Trobe Student Support Service to identify the degree of student satisfaction of the student support services and to identify which service was considered the most important in increasing student satisfaction. In this case study, we only evaluate the satisfaction of the students in the La Trobe Student Support Services based on campus (Albury Wodonga, Bendigo, Bundoora, City, Mildura, Shepparton), as shown in Table 1. The results in Table I show that service satisfaction is higher compare to service popularity

5.3 Evaluate of Social Media Data

In order to identify the participants, we use the sampling pool of Twitter that appears in the NodeXL. It is a software tool that import data from outside data providers. We apply NodeXL, an extendible toolkit for network overview to 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.

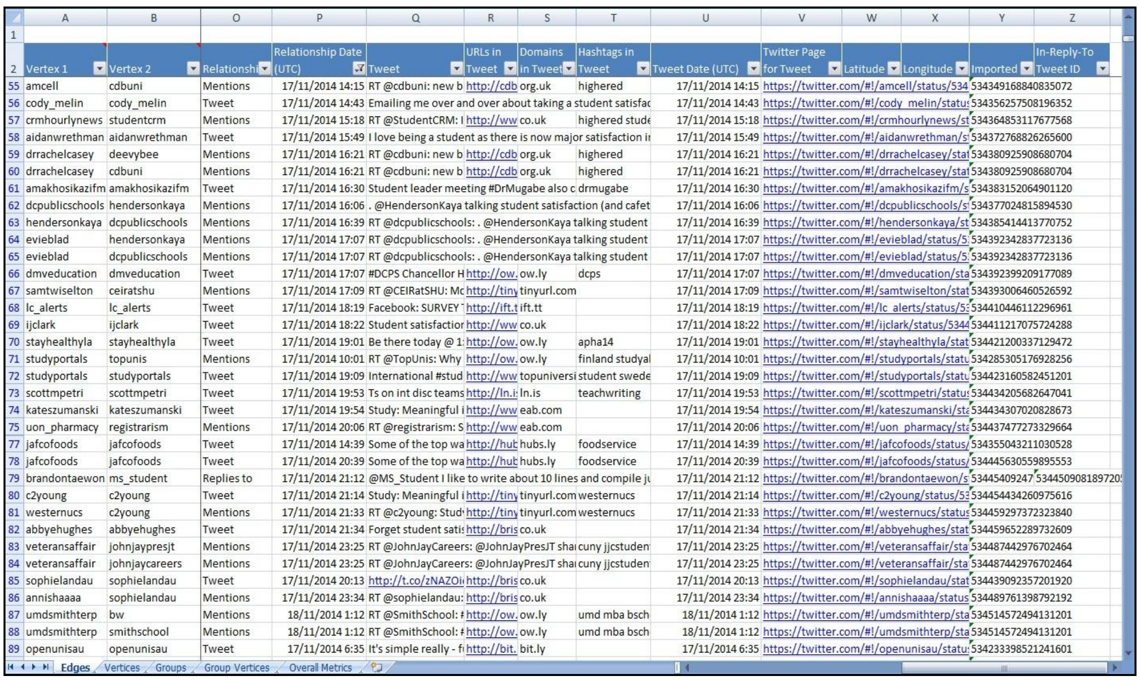
Users create data on Twitter every second and data are collected based on the date of the tweet is created in order to avoid huge volume of data. NodeXL import data from Twitter username LaTrobe into the spreadsheet. In this case study, data are evaluated for two weeks from 17 November 2014 to 30 November 2014, as shown in Table 2. Therefore, data are filter and capture within this period of time.

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. In LaTrobe network, we capture data that match to the query of people who discuss about student satisfaction. We expand this query by looking at people who discuss about service popularity and service satisfaction. Data are based on people who tweet, mention and replies about the query. Data are filter as follows:

- Twitter has many users.
- We capture data from LaTrobe username.
- Within LaTrobe network, we capture data from people who tweet about student satisfaction.
- Within student satisfaction, we capture data from people who tweet about service popularity and service satisfaction.

5.4 Data Filtering and Retrieval Using NodeXL

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.



	A	B	O	P	Q	R	S	T	U	V	W	X	Y	Z
1														
2	Vertex 1	Vertex 2	Relationshi	Relationship Date (UTC)	Tweet	URLs in Tweet	Domains in Tweet	Hashtags in Tweet	Tweet Date (UTC)	Twitter Page for Tweet	Latitude	Longitude	Imported	In-Reply-To Tweet ID
55	amcell	cdbuni	Mentions	17/11/2014 14:15	RT @cdbuni: new b	http://cdb	org.uk	highered	17/11/2014 14:15	https://twitter.com/#!/amcell/status/534			534349168840835072	
56	cody_melin	cody_melin	Tweet	17/11/2014 14:43	Emailing me over and over about taking a student satisfac				17/11/2014 14:43	https://twitter.com/#!/cody_melin/status			534356257508196352	
57	crmhourlynews	studentcrm	Mentions	17/11/2014 15:18	RT @StudentCRM: I	http://ww	co.uk	highered stude	17/11/2014 15:18	https://twitter.com/#!/crmhourlynews/st			534364853117677568	
58	aidanwrethman	aidanwrethman	Tweet	17/11/2014 15:49	I love being a student as there is now major satisfaction i				17/11/2014 15:49	https://twitter.com/#!/aidanwrethman/s			534372768826265600	
59	drrachelcasey	deevybee	Mentions	17/11/2014 16:21	RT @cdbuni: new b	http://cdb	org.uk	highered	17/11/2014 16:21	https://twitter.com/#!/drrachelcasey/stat			534380925908680704	
60	drrachelcasey	cdbuni	Mentions	17/11/2014 16:21	RT @cdbuni: new b	http://cdb	org.uk	highered	17/11/2014 16:21	https://twitter.com/#!/drrachelcasey/stat			534380925908680704	
61	amakhosikazifm	amakhosikazifm	Tweet	17/11/2014 16:30	Student leader meeting #DrMugabe also c			drmugabe	17/11/2014 16:30	https://twitter.com/#!/amakhosikazifm/s			534383152064901120	
62	dcpublicschools	hendersonkaya	Mentions	17/11/2014 16:06	. @HendersonKaya talking student satisfaction (and cafet				17/11/2014 16:06	https://twitter.com/#!/dcpublicschools/s			534377024815894530	
63	hendersonkaya	dcpublicschools	Mentions	17/11/2014 16:39	RT @dcpublicschools: . @HendersonKaya talking student				17/11/2014 16:39	https://twitter.com/#!/hendersonkaya/st			534385414413770752	
64	evieblad	hendersonkaya	Mentions	17/11/2014 17:07	RT @dcpublicschools: . @HendersonKaya talking student				17/11/2014 17:07	https://twitter.com/#!/evieblad/status/5			534392342837723136	
65	evieblad	dcpublicschools	Mentions	17/11/2014 17:07	RT @dcpublicschools: . @HendersonKaya talking student				17/11/2014 17:07	https://twitter.com/#!/evieblad/status/5			534392342837723136	
66	dmveducation	dmveducation	Tweet	17/11/2014 17:07	#DCPS Chancellor H	http://ow.	ow.ly	dcps	17/11/2014 17:07	https://twitter.com/#!/dmveducation/sta			534392399209177089	
67	samtwiselton	ceiratshu	Mentions	17/11/2014 17:09	RT @CEIRatSHU: Mc	http://tiny	tinyurl.com		17/11/2014 17:09	https://twitter.com/#!/samtwiselton/stat			534393006460526592	
68	lc_alerts	lc_alerts	Tweet	17/11/2014 18:19	Facebook: SURVEY	http://ift.t	ift.tt		17/11/2014 18:19	https://twitter.com/#!/lc_alerts/status/5			534410446112296961	
69	ijclark	ijclark	Tweet	17/11/2014 18:22	Student satisfactior	http://ww	co.uk		17/11/2014 18:22	https://twitter.com/#!/ijclark/status/5344			534411217075724288	
70	stayhealthyla	stayhealthyla	Tweet	17/11/2014 19:01	Be there today @ 1	http://ow.	ow.ly	apha14	17/11/2014 19:01	https://twitter.com/#!/stayhealthyla/stat			534421200337129472	
71	studyportals	topunis	Mentions	17/11/2014 10:01	RT @TopUnis: Why	http://ow.	ow.ly	finland studyal	17/11/2014 10:01	https://twitter.com/#!/studyportals/statu			534285305176928256	
72	studyportals	studyportals	Tweet	17/11/2014 19:09	International #stud	http://ww	topuniversi	student swede	17/11/2014 19:09	https://twitter.com/#!/studyportals/statu			534423160582451201	
73	scottmpetri	scottmpetri	Tweet	17/11/2014 19:53	Ts on int disc teams	http://ln.i	ln.is	teachwriting	17/11/2014 19:53	https://twitter.com/#!/scottmpetri/status			534434205682647041	
74	kateszumanski	kateszumanski	Tweet	17/11/2014 19:54	Study: Meaningful i	http://ww	eab.com		17/11/2014 19:54	https://twitter.com/#!/kateszumanski/sta			534434307020828673	
75	uon_pharmacy	registrarism	Mentions	17/11/2014 20:06	RT @registrarism: S	http://ww	eab.com		17/11/2014 20:06	https://twitter.com/#!/uon_pharmacy/sta			534437477273329664	
77	jafcofoods	jafcofoods	Tweet	17/11/2014 14:39	Some of the top wa	http://hub	hubs.ly	foodservice	17/11/2014 14:39	https://twitter.com/#!/jafcofoods/status/			534355043211030528	
78	jafcofoods	jafcofoods	Tweet	17/11/2014 20:39	Some of the top wa	http://hub	hubs.ly	foodservice	17/11/2014 20:39	https://twitter.com/#!/jafcofoods/status/			534445630559895553	
79	brandontaewon	ms_student	Replies to	17/11/2014 21:12	@MS_Student I like to write about 10 lines and compile ju				17/11/2014 21:12	https://twitter.com/#!/brandontaewon/s			53445409247	53445090818972
80	c2young	c2young	Tweet	17/11/2014 21:14	Study: Meaningful i	http://tiny	tinyurl.com	westernucs	17/11/2014 21:14	https://twitter.com/#!/c2young/status/5			534454434260975616	
81	westernucs	c2young	Mentions	17/11/2014 21:33	RT @c2young: Stud	http://tiny	tinyurl.com	westernucs	17/11/2014 21:33	https://twitter.com/#!/westernucs/status			534459297372323840	
82	abbyehughes	abbyehughes	Tweet	17/11/2014 21:34	Forget student sati	http://bris	co.uk		17/11/2014 21:34	https://twitter.com/#!/abbyehughes/stat			534459652289732609	
83	veteransaffair	johnjaypresjt	Mentions	17/11/2014 23:25	RT @JohnJayCareers: @JohnJayPresJT sha			cuny jjcstuden	17/11/2014 23:25	https://twitter.com/#!/veteransaffair/sta			534487442976702464	
84	veteransaffair	johnjaycareers	Mentions	17/11/2014 23:25	RT @JohnJayCareers: @JohnJayPresJT sha			cuny jjcstuden	17/11/2014 23:25	https://twitter.com/#!/veteransaffair/sta			534487442976702464	
85	sophielandau	sophielandau	Tweet	17/11/2014 20:13	http://t.co/zNAZOi	http://bris	co.uk		17/11/2014 20:13	https://twitter.com/#!/sophielandau/stat			534439092357201920	
86	annishaaaa	sophielandau	Mentions	17/11/2014 23:34	RT @sophielandau:	http://bris	co.uk		17/11/2014 23:34	https://twitter.com/#!/annishaaaa/status			534489761398792192	
87	umdsmithterp	bw	Mentions	18/11/2014 1:12	RT @SmithSchool:	http://ow.	ow.ly	umd mba bsch	18/11/2014 1:12	https://twitter.com/#!/umdsmithterp/sta			534514572494131201	
88	umdsmithterp	smithschool	Mentions	18/11/2014 1:12	RT @SmithSchool:	http://ow.	ow.ly	umd mba bsch	18/11/2014 1:12	https://twitter.com/#!/umdsmithterp/sta			534514572494131201	
89	openunisau	openunisau	Tweet	17/11/2014 6:35	It's simple really - f	http://bit.	bit.ly		17/11/2014 6:35	https://twitter.com/#!/openunisau/statu			534233398521241601	

Figure 3. Imported Twitter data into spreadsheet.

Import from Twitter users network.

- It optionally clears the NodeXL workbook, then get the network of specified Twitter users.

- Specify the Twitter users with specific username.
- We are interested in username @LaTrobe.
- Import basic network plus followers and following who replies, mentions and tweet.
- Limit it to 100 recent tweets per user.
- Import from Twitter search network.
- It optionally clears the NodeXL workbook, then get the network of people who tweets certain specified word.
- Search for the tweets that match to the specific query.
- We search for student satisfaction.

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

- Limit to 100 tweets.
- Filter by relationships (tweet, mentions and replies).
- Filter by specific date (day, week, month).
- We apply the steps for tweet that match to service popularity and service satisfaction in user network.

The results in Table 2 show that service satisfaction has the higher number of tweet with 70 numbers. It shows that students prefer to discuss about service satisfaction.

Table 2.score of tweets that relate to student satisfaction.

Weeks	Service popularity	Rank	Service satisfaction	Rank
Week 1	121	1	726	1
Week 2	120	2	524	2
Total/Total rank	241	361	1250	1774
Percentage	67		70	

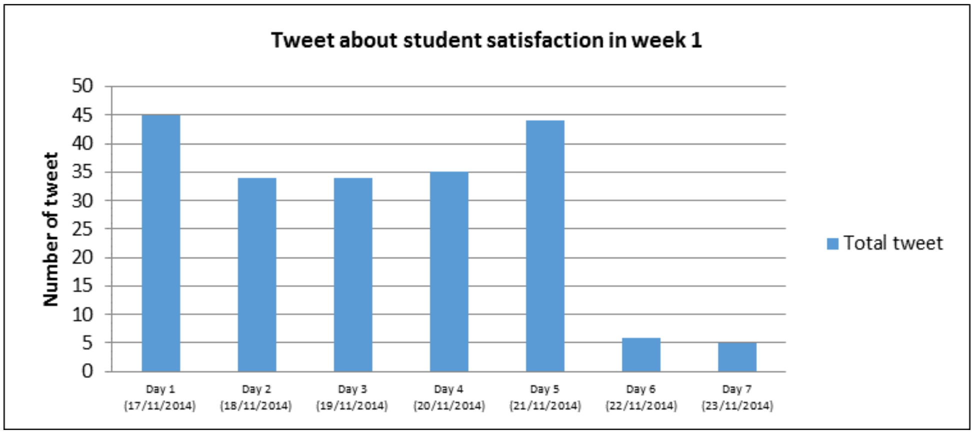


Figure 4. Imported Twitter data into spreadsheet.

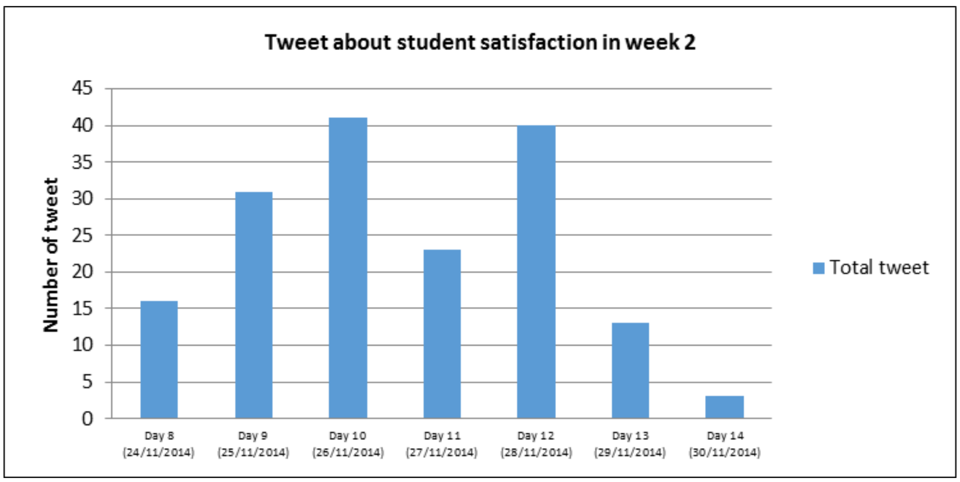


Figure 5. Imported Twitter data into spreadsheet.

5.5 Evaluate of Social Media Data and Organizational Data:

In order to evaluate social data and organizational data, we improve the metrics to incorporate both data in relation to the goal. In this metrics, we evaluate the data based on the overall rank. Overall rank is a total rank for both sub-goals. For example, if total rank for service popularity is 125 and service satisfaction is 130. Therefore, overall rank is 255. The definition of overall rank is as follows:

Overall rank= (sub-goal 1(total rank) + sub-goal 2(total rank))

After we identify the goal and sub-goals, we identify the relationship between goal and sub-goals. Based on this relationship, we identify data that relate to the goal. In this paper, we incorporate internal data (survey from La Trobe University Student Support Service Report) and external data (social data from Twitter).

Table 3.evaluation of social data and organizational data in relation to the student satisfaction.

Data	Service popularity	Rank	Service satisfaction	Rank
Social data	67	1	70	1
Organizational data	29	2	30	2
Total/Total rank	96	125	100	130
Total/ Overall rank	96	255	100	255
Percentage	38		39	

In Table 3, the results show that service satisfaction has the highest percentage compare to service popularity. These results are consistent with the results from organizational data in Table 1 and the results from social data in Table II. The results show that service satisfaction is important toward student satisfaction of La Trobe Student Support Services.

The results show the relationship between social data and organizational data in relation to the case study goal. In conclusion, social data can contribute for better decision-making process.

6. Discussion

Most organizations today are fundamentally dependent on their data and information handling services facilitated by their information technology to collect, store, flow, manage and analyze data better. This paper addressed the relationship between different types of data and organizational goals. We proposed the organizational goals ontology to develop this relationship so we can identify the goals, sub-goals and data that relate to the organizational goals.

A unique contribution of this paper is its perspective on how data from social media can support organizational data for better decision-making. The paper resolves the issue to identify data that relate to the goals, especially social data. After we analyse both organizational data and social data, we see that these two data provide consistent results. The results assist decision-making process for consistent outcome, which may lead to achieving the organizational goals.

Evidence has shown that organizational goals ontology can be effectively developed the relationship between organizational data and social data in relation to the organizational goals. 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 they could incorporate with organizational data for better decision-making. Also, as larger 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.

6.1 Contribution

The analysis and evaluation of the data assisted our decision-making process in evaluating the level of student satisfaction based on service satisfaction and service popularity based on organizational data and social data. The case study was implemented and proves the applicability of the organizational goals ontology. In contrast to Rao et al. (2012), Fox et al. (1998) and Sharma and Osei-Bryson (2008), the final results of the organizational goals ontology achieved the aim of being flexible and applicable in assisting decision-making in relation to the organizational goals.

- *Flexible to identify the organizational goals*

In this paper, we explained how to define the organizational goals. The usage of ontology assists the flexibility to define the organizational goals. By using an ontology, the process to identify the set of the organizational goals becomes flexible. The results in the case study proved the flexibility how we want to define the main goal.

- *Flexible to identify the dependency relationship*

The organizational goals ontology develops the dependency relationship between organizational data and organizational goals. We developed the relationship between social data and organizational goals. We explained how to identify organizational data from organizational datasets that relate to the organizational goals. The organizational goals ontology provides this flexibility to develop this dependency. We proved this flexibility in which we developed the dependency relationship between data (internal and external data) and case study goals. This flexibility assists the process to identify which data to be consider relevant to the organizational goals.

- *Flexible to define the metrics after the main goals are identified*

We then test the flexibility to define the metrics. In this paper, the organizational goals ontology gives domain experts and entrepreneurs the flexibility on how they want to define the metrics after they identified the main goals. Domain experts and entrepreneurs have this flexibility on how they want to evaluate the data that relate to the organizational goals. This flexibility was tested in the case study. This proves that the organizational goals ontology assists the process to define the metrics in different way after we identified the goals that we want to evaluate.

- *Organizational goals ontology assists the decision-making and provide feedback in relation to the organizational goals*

The main objective of data analysis is to evaluate data from the vast amount of datasets. In this paper, data analysis is important to identify the value of data that relevant to the goals to support decision-making process in relation to the organizational goals (Izhar et al., 2013).

After we analyse the data based on the metrics, the values of this analysis were presented to evaluate the level of the organizational goals achievement. Organizational goals ontology aims to provide a platform to analyse this value. Therefore, they can evaluate the level of the organizational goals achievement.

7. Future work

In this paper, the evaluation of external data only limited to social media. There are other external data that can support the internal data for efficient decision-making such as mobile data and sensor data. Therefore, it is important to evaluate this data to see if this data can be used for decision-making process.

In the future, we will take this framework to the next level by building a Global Ontology that can capture the concepts of complexity and use as a basis for data integration processes suitable for efficient correlative decision-making. This Global Ontology will reside on the cloud and it will initially contain the local organizational ontology.

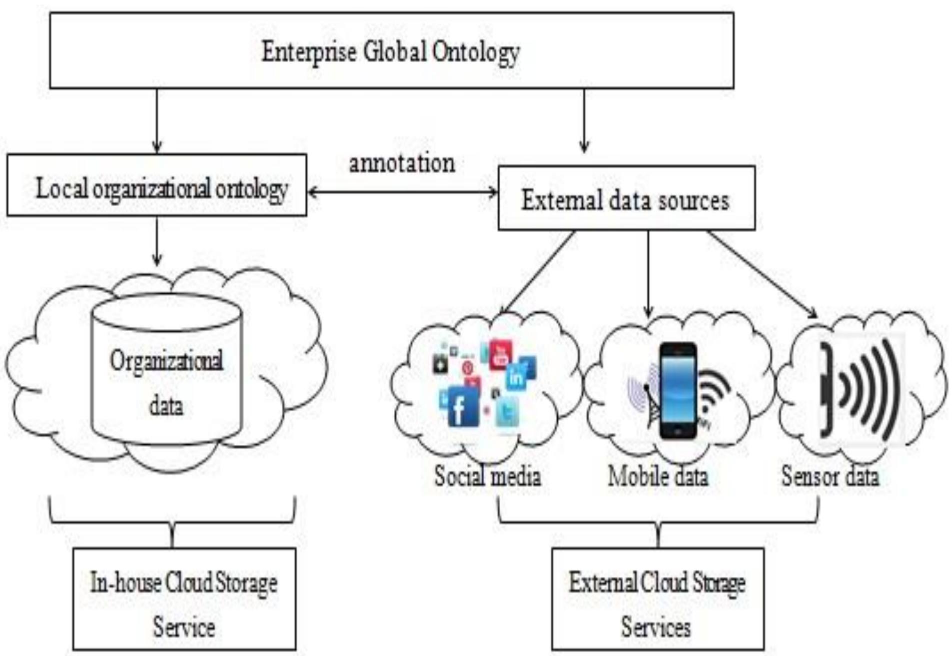


Figure 6. Future work.

We will develop an incremental approach to extend the ontology with relevant external knowledge bases by clustering and annotating the external data to concepts and properties within the ontology.

8. Conclusion

We have described the main features of the organizational goals ontology when developed the relationship between organizational data and social data in relation to the organizational goals. In addition, we have proposed an alternative approach to capture social data using NodeXL. We have tested the application of the organizational goals ontology for better decision-making in relation to the organizational goals by evaluating both organizational data and social data. In conclusion, organizations can rely on external data such as social data to support their decision-making to evaluate their organizational goals achievement. This paper shows that the evaluation of the organizational goals achievement is better by incorporating social data and organizational data.

References

Almeida, M. B., & Barbosa, R. R. (2009). Ontologies in knowledge management support: a case study. *Journal of the American Society for Information Science and Technology, 60*(10), 2032-2047.

Assuncao, M. D., Calheiros, R. N., Bianchi, S., & Netto, M. A. S. (2015). Big data computing and clouds: trends and future directions. *Journal of Parallel and Distributed Computing, 79*, 3-15.

Berber, M., Graupner, E., & Maedche, A. (2014). The information panopticon in the big data era. *Journal of Organization Design, 3*(1), 14-19.

Cho, J., Han, S., & Kim, H. (2006). Meta-ontology for automated information integration of parts libraries. *Computer-Aided Design, 38*(7), 713-725.

Davenport, T. H., & Dyche, J. (2013). Big data in big companies: International Institute for Analytics.

Fox, M. S., Barbuceanu, M., & Gruninger, M. (1996). An organisation ontology for enterprise modeling: Preliminary concepts for linking structure and behaviour. *Computers in Industry, 29*(1-2), 123-134.

Fox, M. S., Barbuceanu, M., Gruninger, M., & Lin, J. (1998). An organization ontology for enterprise modelling *Simulation organizations: Computational models of institutions and groupsAAAI/MIT Press* (pp. 131-152).

Galbraith, J. R. (2014). Organization design challengers resulting from big data. *Journal of Organization Design, 3*(1), 2-13.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods and anlytics. *International Journal of Information Management, 35*, 137-144.

Gangadharbatla, H., Bright, L. F., & Logan, K. (2014). Social media and news gathering: tapping into the millennial mindset. *The Journal of Social Media in Society, 3*(1), 45-63.

Governatori, G., & Iannella, R. (2011). A modeling and reasoning framework for social networks policies. *Enterprise Information Systems, 5*(1), 145–167.

Grossman, R. L., & Siegel, K. P. (2014). Organizational models for big data and analytics. *Journal of Organization Design, 3*(1), 20-25.

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics and big data in supply chain management: An introduction to the problem and suggestions for research and application. *International Journal Production Economics, 154*, 72-80.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in pizza industry. *International Journal of Information Management*
33, 464-472.

Huang, T., Lan, L., Fang, X., An, P., Min, J., & Wang, F. (2015). Promises and challenges of big data computing in health sciences. *Big Data Research, 2*(1), 2-11.

Izhar, T. A. T., Torabi, T., Bhatti, M. I., & Liu, F. (2013). Recent developments in the organization goals conformance using ontology. *Expert Systems with Applications, 40*(10), 4252-4267.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research, 2*, 59-64.

Kang, J., & Sim, K. M. (2011). *Ontology and search engine for cloud computing system*. Paper presented at the IEEE 2011 International Conference on System Science and Engineering (ICSSE).

Kaplan, A. M., & Haelein, M. (2010). Users of the worl, united! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59-68.

Keckley, P. H. (2010). Social networks in health care: Communication, collaboration and insights. from http://www.deloitte.com/ assets/Dcom-UnitedStates/Local%20Assets/Documents/US CHS 2010Social Networks 070710.pdf

Mansingh, G., Osei-Bryson, K.-M., & Reichgelt, H. (2009). Building ontology-based knowledge maps to assist knowledge process outsourcing decisions. *Knowledge Management Research and Practice, 7*, 37-51.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition and productivity. *McKinsey Global Institute* from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

McIntyre, K. (2014). The evolution of social media from 1969 to 2013: A change in competition and a trend toward complementary. *The Journal of Social Media in Society, 3*(2), 6-25.

Park, N., Kee, K. F., & Valenzuela, S. (2009). Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes
CyberPsychology & Behavior, 12(6), 729–733.

Polato, I., Re, R., Goldman, A., & Kon, F. (2014). A comprehensive view of Hadoop research- A systematic literature review. *Journal of Networking and Computer Application*

46, 1-25.

Qualman, E. (2009). *Socialnomics how social media transforms the way we live and do business*. Hoboken: Wiley John & Sons, Inc.

Raacke, J., & Bonds-Raacke, J. (2008). MySpace and Facebook: Applying the uses and gratifications theory to exploring friend-networking sites. *CyberPsychology & Behavior, 11*(2), 169–174.

Rao, L., Mansingh, G., & Osei-Bryson, K.-M. (2012). Building ontology based knowledge maps to assist business process re-engineering. *Decision Support Systems, 52*(3), 577-589.

Rao, L., Reichgelt, H., & Osei-Bryson, K.-M. (2009). An approach for ontology development and assessment using a quality framework. *Knowledge Management Research and Practice, 7*(3), 260-276.

Sá, J. O. e., Martins, C., & Simões, P. (2015). Big Data in Cloud: A Data Architecture. *New Contributions in Information Systems and Technologies, 353*, 723-732.

Safko, L., & Brake, D. K. (2009). *The social media bible: Tactics, tools, and strategies for business success*. Hoboken: Wiley John & Sons, Inc. .

Sharma, S., & Osei-Bryson, K.-M. (2008). *Organization-ontology based framework for implementing the business understanding phase of data mining projects.* Paper presented at the International Conference on System Sciences, Hawaii.

Silva, N. F. F. d., Hruschka, E. R., & Jr, E. R. H. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems, 66*, 170-179.

Valaski, J., Malucelli, A., & Reinehr, S. (2012). Ontologies application in organizational learning: A literature review. *Expert Systems with Applications, 39*(8), 7555-7561.

Valiente, M.-C., Garcia-Barriocanal, E., & Sicilia, M.-A. (2012). Applying an ontology approach to IT service management for business-IT integration. *Knowledge-Based Systems, 28*, 76-87.

Zhi, Q., Zhao-Wen, L., & Yan, M. (2011). Research of Hadoop-based data flow management system. *The Journal of China Universities of Post and Telecommunications, 18*, 164-168.