ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: <http://www.elsevier.com/locate/kontakt>

Original research article

Use of the Omaha System for ontology-based text mining to discover meaning within CaringBridge social media journals

Karen A. Monsen ^{a*}, Sasank Maganti ^b, Robert A. Giaquinto ^b, Michelle A. Mathiason ^a,
Ragnhildur I. Bjarnadottir ^c, Mary Jo Kreitzer ^d

^a University of Minnesota, School of Nursing, Minneapolis, USA

^b University of Minnesota, Computer Science and Engineering, Minneapolis, USA

^c University of Florida, College of Nursing, Gainesville, Florida, USA

^d University of Minnesota, Earl E. Bakken Center for Spirituality and Healing, Minneapolis, USA

ARTICLE INFO

Received: 2017-11-15

Received in revised form:

2018-01-31

Accepted: 2018-03-06

Published online: 2018-10-12

Keywords:

Text mining

Social media

Ontology

Omaha System

CaringBridge

Terminology

ABSTRACT

Objectives: The goals of this study were to examine the feasibility of using ontology-based text mining with CaringBridge social media journal entries in order to understand journal content from a whole-person perspective. Specific aims were to describe Omaha System problem concept frequencies in the journal entries over a four-step process overall, and relative to Omaha System Domains; and to examine the four step method including the use of standardized terms and related words.

Design: Ontology-based retrospective observational feasibility study using text mining methods.

Sample: A corpus of social media text consisting of 13,757,900 CaringBridge journal entries from June 2006 to June 2016.

Measures: The Omaha System terms, including problems and signs/symptoms, were used as the foundational lexicon for this study. Development of an extended lexicon with related words for each problem concept expanded the semantics-powered data analytics approach to reflect consumer word choices.

Results: All Omaha System problem concepts were identified in the journal entries, with consistent representation across domains. The approach was most successful when common words were used to represent clinical terms. Preliminary validation of journal examples showed appropriate representation of the problem concepts.

Conclusions: This is the first study to evaluate the feasibility of using an interface terminology and ontology (the Omaha System) as a text mining information model. Further research is needed to systematically validate these findings, refine the process as needed to advance the study of CaringBridge content, and extend the use of this method to other consumer-generated journal entries and terminologies.

© 2018 Jihočeská univerzita v Českých Budějovicích, Zdravotně sociální fakulta.
Published by Elsevier Sp. z o. o. All rights reserved.

* **Author for correspondence:** Karen A. Monsen, PhD, RN, FAAN, University of Minnesota, School of Nursing, 5-140 Weaver-Densford Hall, 308 Harvard Street SE, Minneapolis, MN 55455, USA; e-mail: mons0122@umn.edu
<http://dx.doi.org/10.1016/j.kontakt.2018.03.002>

KONTAKT XX/3: 227–233 • ISSN 1212-4117 (Print) • ISSN 1804-7122 (Online)

Please cite the article Monsen KA, Maganti S, Giaquinto RA, Mathiason MA, Bjarnadottir RI, Kreitzer MJ. Use of the Omaha System for ontology-based text mining to discover meaning within CaringBridge social media journals. *Kontakt* 2018; 20(3): e210–e216; <http://dx.doi.org/10.1016/j.kontakt.2018.03.002>

Introduction

Social network technologies are mainstream communication tools for social support and information sharing. Numerous studies have shown that virtual support networks are valuable tools for those suffering from chronic conditions or life-threatening illness and their caregivers [1–3]. There is a need for health care clinicians and researchers to understand content within social networks to aid in understanding consumers' perspectives during a health crisis [4]. CaringBridge is a web-based social network used widely for social support, often during a health crisis that specifically promotes wellbeing within the context of a caring community [5,6].

Text mining for understanding content of CaringBridge journals

Text mining is a potential strategy for understanding content within compassionate technologies and social network communities. However, the extraction and analysis of large volumes of narrative text poses significant challenges. Because of the variability and complexities of language, the data cannot be analyzed using traditional statistical algorithms [7].

Text mining methods

The healthcare domain has seen an explosion in applications of text mining to inform numerous health, business, and social research agendas [1–5, 8–16]. For example, Wen et al. [14] showed that social support networks contain rich data, and data mining approaches used to study quantitative approaches can be used to study cancer support communities. Semantic-powered approaches rely on existing domain knowledge and have been widely used in linguistic research and health informatics research [15,16]. However, semantic-powered approaches have not previously been applied to social media text corpora using a standardized interface terminology.

Interface terminologies as domain knowledge models

Among the tools used for health care knowledge representation, interface terminologies are designed for use at the human-computer interface [17, 18]. Because interface terminologies describe health and social needs and interventions, we were interested in exploring the application of an interface terminology to journal entries in social media as a potentially useful perspective for text mining in social media data. There are 12 standardized terminologies recognized by the American Nurses Association [17]. Of these, the Omaha System was designed to be suitable for use across disciplines (including use with consumers). It is a psychometrically sound instrument and comprehensive, holistic ontology for health that is particularly well-suited to use in ontology-based text mining because of its robust taxonomic, hierarchical, multi-axial structure [18–21]. The Omaha System classifies

health as a comprehensive whole using 42 concepts (called problems) in four domains: Environmental (4 concepts), Psychosocial (12), Physiological (18), and Health-related Behaviours (8), (Chart 1).

The purpose of this study was to examine the feasibility of using an ontology-based structured text mining approach based on the Omaha System with CaringBridge social media journal text data. The long term goals of our research are to discover meaning within clinically-relevant text. For example, we seek to discover patterns that may suggest intervention approaches, and to understand outcomes such as those that may be expressed in CaringBridge journal entries. Specific aims were to describe Omaha System problem concept frequencies in the journals over the four-step process overall in order to evaluate the feasibility of using the Omaha System at each step; evaluate the results relative to the four Omaha System Domains; and to examine the content of the Caring Bridge corpus across all 42 problem concepts.

Materials and methods

Setting

This feasibility study was conducted using de-identified CaringBridge journal entries in accordance with CaringBridge privacy policies and with permission of the CaringBridge leadership; and was deemed exempt from IRB review by the University Institutional Review Board. Established in 1997, CaringBridge is a web-based social network serving over 500,000 people daily and over 43 million people annually in over 230 countries. Within the CaringBridge social media platform, personal, protected websites are created by individuals or their caregivers to share information and photos about their health journey and provide updates to a community of friends who are able to offer support and encouragement. CaringBridge is a 501(c) (3) non-profit organization whose overall objective is to use technology to enhance social connectedness, and bring people together to help overcome the isolation that comes with illness. The majority of CaringBridge sites are created for people with serious and life threatening illnesses. The interdisciplinary research team (CaringBridge Research Collaborative) is comprised of scientists and graduate students from the University. The data were stored in the University's secure Supercomputing Institute.

Study design and sample

This retrospective, observational text mining study reused archival CaringBridge data (June 1, 2005–June 3, 2016) that was de-identified prior to transferring and storage within the Supercomputing Institute. There were 13,757,900 journal entries in the corpus.

Text mining methods

The semantics-powered text mining approach developed for this study is described below. While similar approaches

Problem Concept by Domain	Journal Entries in Combined Stems (Step 2)	Related Words	Journal Entries after Related Words (Step 4)	Fold increase
Environmental Domain				
Sanitation	995	14	72,448	91.91
Income	42,954	16	368,604	7.58
Neighborhood/workplace safety	182,132	21	1,061,596	4.83
Residence	31,545	16	4,265,756	134.23
Psychosocial Domain				
Role change	336	3	3,453	9.28
Abuse	14,695	9	14,814	0.01
Social contact	351	15	25,186	70.75
Caretaking/parenting	5,158	10	25,765	4
Neglect	28,609	5	32,758	0.15
Sexuality	980	17	59,379	59.59
Grief	65,112	5	159,876	1.46
Growth and development	322,655	13	1,411,400	3.37
Mental health	10,256	45	1,575,419	152.61
Interpersonal relationship	111,535	21	1,577,683	13.15
Communication with community resources	294,682	26	2,213,677	6.51
Spirituality	3,801	19	5,889,697	1548.51
Physiological Domain				
Postpartum	1,221	13	32,029	25.23
Consciousness	18,686	8	103,552	4.54
Cognition	30,100	17	169,237	4.62
Pregnancy	34,331	12	252,202	6.35
Reproductive function	2,374	17	373,140	156.18
Bowel	102,008	31	500,027	3.9
Urinary function	26,872	9	771,223	27.7
Oral health	288,226	18	1,054,955	2.66
Skin	384,459	23	1,136,802	1.96
Vision	257,013	24	1,425,489	4.55
Digestion-hydration	92,870	23	1,762,650	17.98
Respiration	15,710	23	1,841,676	116.23
Hearing	92,595	18	1,847,297	18.95
Speech and language	377,959	15	2,984,742	77.01
Pain	2,180,946	10	3,278,011	0.5
Circulation	21,342	38	3,691,933	171.99
Neuro-musculo-skeletal function	352,501	52	5,855,950	15.61
Communicable/infectious condition	62,952	71	5,898,949	92.71
Health-related Behaviors Domain				
Family planning	913	31	7,838	7.58
Substance use	359	22	67,314	186.5
Health care supervision	58,158	14	245,833	3.23
Personal care	6,791	16	461,673	66.98
Physical activity	12,758	22	725,523	55.87
Medication regimen	758,510	26	3,259,890	3.3
Nutrition	183,293	32	3,867,513	20.1
Sleep and rest patterns	4,251,293	13	4,599,111	0.08

Chart 1 – Numbers of journal entries, related words, and fold increases for Omaha System problem concepts by domain

have been used in previous health sciences research, it is novel in the application of the Omaha System ontology and the focus on wellbeing and compassionate technologies. This method examined the ability of the Omaha System terms and related words to tag and retrieve journals with problem concept-related content. To fully explore all Omaha System problem concepts within the corpus, a list of problem-specific related words for each of the Omaha System problem concepts was developed collaboratively based on domain expertise.

Data preprocessing

We preprocessed the corpus using standard preprocessing methods [22]. First excess whitespace and non-text characters were removed. Next, hyphenated words and words combined with numbers were split, and contractions were expanded. Finally, we used part-of-speech tagging and the Wordnet Lemmatizer to standardize words to their simplest form (e.g. “resting”, “rested”, and “rest” are all transformed to “rest”).

Steps in the text mining process

The text retrieval and query expansion steps were repeated for the 42 Omaha System problem concepts using Python 2.7.

- Step 1: For each term or stem of the problem concept, look for problem terms and stems in the entire corpus. Each journal containing a term or stem was retrieved and tagged.
- Step 2: Journal entries containing the terms and stems for the same problem concept were tagged for the respective problem concept name.
- Step 3: From the remaining corpus, journals with additional problem-specific signs/symptom terms and stems were retrieved and tagged.
- Step 4: From the remaining corpus, journals with problem-specific related words were retrieved and tagged.

Journal entry validation

Selected sentences were abstracted to illustrate journal content labelled with Omaha System problem concepts based on the text mining approach. These labels were evaluated by content experts (authors) for face validity.

Data analysis

Standard descriptive and inferential statistics were employed to evaluate problem concept frequencies in the journals over the four-step process overall, and after step 4, relative to Omaha System Domains.

Results

The overall results are depicted in Fig. 1. A detailed example of the semantic-powered text mining multi-step process is provided for the problem concept of Sleep and Rest Patterns, defined as “Periods of suspended motor and sensory activity and periods of inactivity, repose, or mental calm” [18]. For the Sleep and rest problem there are eight signs/symptoms: “sleep/rest pattern disrupts family, frequently wakes during night, sleepwalking, insomnia, nightmares, insufficient sleep/rest for age/physical condition, sleep apnea, and snoring” [18].

In step 1, the terms and stems at the problem level were Sleep (2,275,504), Rest (2,685,494). In step 2, the two term combination using the conjunction “or”: i.e. Sleep or Rest (4,251,293). In step 3, signs/symptoms terms and stems were awake during night, or sleepwalk, or insomnia, or nightmare, or snoring; combined with Sleep or Rest (4,280,625). In step 4, the results of step 3 were combined with the related words repose, or snore, or awake at night, or awake all night, or up all night, or nap, or doze, or shut eye, or shuteye (4,599,111).

Feasibility of using the Omaha System

Aim 1 was to describe Omaha System problem concept frequencies in the journals over the four-step process overall, and relative to Omaha System Domains. In step 1, stems ($N = 54$) for all Omaha System problem concepts were present in journal entries. Frequencies of the stems ranged from 336 (Role change) to 2,685,494 (Rest). In step 2, combining the 54 stems into 42 problems resulted in the reordering of problem concepts, and a new maximum of 4,251,293 journals for the Sleep and rest patterns problem concept. At this stage, each problem concept was found on average in 1.86% of journals. In step 3, terms and stems derived from signs/symptoms and definitions were added, showing incremental increases across all steps for most problems. A larger increase occurred in step 4, after adding related words derived from consumer versions of the Omaha System, clinical and terminology expertise, and internet searches, resulting in <1% increase (Abuse) to 1548-fold increase (Spirituality).

After step 4, the Omaha System problem concepts with the highest frequency were Communicable/infectious condition, Spirituality, and Neuro-musculo-skeletal function, all found in more than 5.8 million journals (more than

42% of journals); followed by Sleep and rest patterns and Residence, found in more than 4.2 million journals (more than 31% of journals). Of the 42 Omaha System problem concepts, 21 (50%) were found in one million or more journals. On average, any problem concept may be found in 11.24% of all journals [range = 0.03% (Role change) to 42.88% (Communicable/infectious condition)].

Results by Omaha System Domain showed that the five Omaha System problem concepts with the highest frequencies were from across all four Omaha System Domains (Communicable/infectious condition [Physiological], Spirituality [Psychosocial], Neuro-musculo-skeletal function [Physiological], Sleep and rest patterns [Health-related Behaviours], and Residence [Environmental] (Fig 1). On average, problem concept counts did not differ in the number of journals per problem by domain, ranging from 7.87% (Psychosocial) to 13.32% (Physiological). We also tested differences by domain in the number of journals per problem concept at step 2, percentage increase from step 2 to step 4, and number of related words. The one-way ANOVA showed no significant difference between domains overall (step 1 $p = 0.324$, step 2 $p = 0.740$, step 3 $p = 0.592$, step 4 $p = 0.394$), nor was there significant heterogeneity (Tukey *post hoc* analysis, step 1 $p = 0.375$, step 2 $p = 0.849$, step 3 $p = 0.772$, step 4 $p = 0.637$).

Number of related words

Aim 2 was to examine the use of standardized problem terms and stems, signs/symptoms terms and stems, and related words in the CaringBridge corpus. We tested the relationship between the counts of related words added to the 42 terms/stems to determine whether more related words contributed to larger increases. The average number of related words was 20.3 (range = 3–71; SD = 12.9). The analysis showed that problem concept-specific word count was not correlated with the size of the increase in total number of journals (Pearson’s correlation = 0.061; $p = 0.699$).

Preliminary validation of Omaha System concepts within journal entries was conducted through review of randomly selected journals. Examples of Journal entries with Problem Concepts Labels in brackets are provided below.

- “He is experiencing a little more nausea [Digestion-hydration] feeling and more diarrhoea [Bowel function]. But those things don’t bring the same concern level as his heart [Circulation].”
- “Of course this was the first night I didn’t go right to sleep [Sleep and rest patterns] (no nausea or pain meds [Digestion/hydration, Pain, Medication regimen]) so I started stewing about my staples becoming part of me.”
- “Please pray [Spirituality] her eyes are healed and the next exam shows retinopathy [Vision]. Let’s face it, she needs great eyesight to help her dad [Caretaking/parenting], pick out all the animals on the trail and see a trout below the surface. Continued improvements in liver function [Digestion-hydration] and blood sugar [Nutrition].”

- “A real distrust [Interpersonal relationship] to any change in environment [Neighbourhood/workplace safety]. You can’t explain things to him anymore cause it all hurts and is uncomfortable [Pain].”
- “Things looked fine as far as the skin [Skin] colour, oozing, swelling, and blisters, it’s all there, just in the right amounts, nothing to worry about.”
- “They will run some tests to see if parainfluenza [Communicable/infectious condition] is still the cause of his cough [Respiration].”
- “He does not walk unless necessary, like getting to the car, so he’s quite weak [Neuro-musculo-skeletal function].”
- “His team includes his doctor, head P.T., O.T, nurse, therapist, case manager, and I think another person but can’t think of who it is [Health care supervision]? We were craving more concrete answers so we knew how to proceed with home modifications [Residence] and school [Communication with community resources] and... well life.”
- “At age 12, my father [Caretaking/parenting] raped me [Abuse], with my mother in the other room, and I kept that secret until I was 30 years old. I finally couldn’t hold in my pain anymore [Pain]. Like most abused girls [Growth and development], I turned to indiscriminate sex [Sexuality], drugs [Substance use] and daily demeaning activities that I felt I was selling my soul to partake in. This story is the abbreviated start to what eventually became a very happy [Mental health] ending.”

Discussion

This study examined the feasibility of using an ontology-based text mining approach with CaringBridge social media journal entries using the Omaha System. All Omaha System problem concepts were identified in the corpus of CaringBridge author journal entries using Omaha System terms. Adding related words to the original terms increased the success of the approach and changed relative frequencies of the distribution. Examples of journals show appropriate representation of the problem concepts. This is the first known study using the Omaha System interface terminology for semantics-powered data analytics in social media data. Further research is needed to systematically validate these findings and refine the process as needed to advance the study of CaringBridge content, and extend the use of this method to other consumer- and clinician-generated narratives.

The findings of Aim 1, describing Omaha System problem concept frequencies in the journals over the four-step process, showed that Omaha System terms were found frequently in CaringBridge journals. This aligns with the intent of the Omaha System from its inception to serve a simple, consumer-informed, multidisciplinary terminology [18]. Consistent with our previous research, the challenge of understanding the whole person perspective within the journal entries was facilitated by the comprehensive, holistic structure of the four Omaha System Domains

[18–20]. Relative to Omaha System Domains, the finding that there was no significant difference between or among the domains is consistent with the multi-domain notion of health from a consumer perspective, as may be anticipated in social media narratives [18–20]. In other words, we may assume that all domains are equally important for the expression of content important to CaringBridge journal authors. This supports the use of the Omaha System as a tool to examine CaringBridge journal meaning and furthermore that CaringBridge journal data may be useful to critique the completeness of the Omaha System from the consumer perspective.

The findings of Aim 2, examining the use of standardized problem terms and stems, signs/symptoms terms and stems, and related words in the CaringBridge corpus, showed that while standardized terms incrementally improved the identification of Omaha System concepts in CaringBridge journals, the related words added in step 4 were a key to identifying the extent to which Omaha System concepts were present. This finding highlights the challenges of using standardized terminologies as the basis for consumer-generated text, and emphasizes the importance of creating a robust lexicon of related words for a standardized term. The fact that step 4 findings differed greatly from step 1 findings highlights the limitations of semantics-powered text mining using standardized terms only. Use of a standardized terminology for text mining may be considerably more challenging in social media data than in a corpus of clinician-generated text. The success of this approach with social media text lies in the identification of related words that accurately capture the meaning of each standardized term.

The finding that more problem concept-related words were not associated with a larger increase in number of journals expressing the concept is surprising, given the abundance of ways in which a single concept may be expressed using natural language. Rather, commonly employed simple words appear to be more important in the text mining approach than many more complicated and granular terms. Further research is needed to identify such words for concepts of interest. In alignment with the Omaha System public domain policy, the complete list of problem concept terms and related words will be made available in the public domain to enable translation between text and structured data, and further research using this robust information model [18]. Next steps include the refinement of the problem-concept lexicon, and the evaluation of Omaha System intervention concept terms.

The finding of the preliminary problem concept validation within journals was promising. Overall the exemplar journal entries showed that appropriate representation of the problem concepts using the ontology-based method was successful. Exceptions include the word “distrust”, which was used to identify the Interpersonal relationship problem concept, and which may or may not fit the overall intent of the CaringBridge author in the sentence. Likewise, the use of the word “environment” to identify the Neighbourhood/workplace safety problem concept may be tangential but related. In the final example, the word “pain” may have been used to

express psychological pain and therefore would be more appropriate for the psychosocial mental health problem concept, rather than the physiological Pain problem concept. Further refinement of the set of related words for each concept will continue as part of the validation of these findings. Nevertheless, these brief examples reveal multiple problem concepts in short excerpts from journal entries, and highlight the richness of the content within the journals. The finding that several problems were found within the same journals demonstrates that the Omaha System is a useful information model to aid in interpreting the rich content of CaringBridge journal entries, and is a promising first step towards pattern discovery across like journals. There is potential to label journals meaningfully to enable clustering and comparison of outcome across groups.

Numerous content-related findings contributed to preliminary understanding about concepts addressed by CaringBridge authors. For example, sleep and rest patterns, pain, and medication regimen problem concepts were prominent in the data at step 1 and remained among the most frequent problem concepts throughout all steps. This is consistent with the immediate needs of individuals who are experiencing life threatening crises, life-altering events, or end of life situations; as they often need increased sleep, and are dealing with pain symptoms and the challenges of managing medications [19]. In addition, numerous disease-specific diagnosis terms and physiological descriptions were identified in the journal entries, as authors reported detailed updates. After further validation, such findings may aid in understanding the range of concepts that are expressed by CaringBridge authors, and may lead to future supportive interventions for CaringBridge users. For example, there is a need to understand the content of social networking in order to make a site such as CaringBridge more valuable to users' and responsive to their needs. Knowing, that sleep and rest are such predominant issues within the CaringBridge community offers CaringBridge the opportunity to develop resources that target this need. Sleep is a very significant issue and the lack of sleep is associated with diminished concentration and ability to manage stress as well as increased morbidity and mortality. A site such as CaringBridge could offer information to users on simple self-care strategies known to be effective such as sleep hygiene measures (go to bed when sleepy, develop sleep rituals, avoid caffeine, alcohol and nicotine 4–6 h before bedtime), mind-body and movement interventions (meditation, yoga and Reiki) and use of natural products [23]. Insomnia related costs range from \$30 to \$35 billion per year to \$92.5 to \$107 billion per year [24]. These costs include increased consumption of medical services, increased accident risk and workplace productivity. If even a small number of CaringBridge users benefited from accessing information on the CaringBridge site that led to improved sleep, the financial impact as well as the impact on quality of life would be substantial.

These findings highlight the utility of the ontology-based text mining approach to identify important content, illuminating challenges faced by CaringBridge authors in the context of whole-person health. Findings support the notion that a whole-person perspective results from the use

of an ontology-based approach and ensures that important concepts are not overlooked [19, 20]. A compelling finding was noted during preprocessing, when a simple word count revealed numerous author expressions of intense emotions (e.g. "aaaaaaagony"). This and the fact that all Omaha System problem concepts were included in the corpus also demonstrated that the social media platform CaringBridge is perceived by users as a safe site to discuss deeply personal and meaningful details of their lives. Such discoveries should be incorporated into processes designed to analyze sentiment within the journals. Furthermore, the CaringBridge corpus has potential for use in adding to or refining the Omaha System as a consumer-facing terminology.

This first study using the Omaha System as an ontological approach to text mining was promising and establishes a precedent for further study, even while there were several limitations. Simple word counts may be useful for understanding topics addressed within journal entries, but the method does not fully describe the way in which the words were used (e.g. positive or negative sentiment) or for whom (the author or another person) the words were used. This is useful in the sense that we know authors are interested in the concept, and therefore may be interested in further information or other intervention related to that concept. Furthermore, text mining studies are subject to bias due to the need to take words at face value, despite the fact that words are often used idiomatically. Therefore in this study, we did not include some words that represented too much ambiguity: e.g. "see" could relate to Vision or be used in many other idiomatic expressions. In social media narratives, there are many ways to say the same thing, and it is difficult to capture all synonyms and related words. Moreover, this study did not address negation: e.g. the search for "pain" would also extract instances of "no pain". Despite the taxonomic structure of the Omaha System, brief journals were found to contain numerous inter-related problems, and there may be areas of overlap between some concepts in natural language. It is challenging to classify natural language for these reasons. Finally, as this was an exploratory study, the precision and recall of the text mining approach was not evaluated. All of these limitations should be addressed during future refinements of the approach with attention to excluding words that may be problematic taxonomically or idiomatically, and using words or phrases discovered through formal validation of journal content relative to problem concepts. Even with these limitations, after step 4, over half of the Omaha System problem concepts were found in at least a million journal entries, demonstrating the importance and usefulness of the concepts and the approach. Future studies of clinician-generated notes using this approach will examine the perspectives of clinicians relative to whole-person wellbeing as demonstrated through CaringBridge findings.

Conclusion

This study established preliminary feasibility of using an ontology-based structured text mining approach with

CaringBridge social media journal entries, and provides a platform for continued research using this approach. This research extends from classification to data abstraction and is an exploratory step toward outcomes evaluation using text data from social media and other sources such as clinical notes based on the structure of the Omaha System. Further research is needed to validate findings, to refine lexicons for Omaha System problems, and to extend the use of this method to other text data.

Conflict of interests

The authors declare no conflict of interests in regards to this article.

Acknowledgements

Financial support for this research was provided by CaringBridge, a non-profit located in Eagan, Minnesota.

REFERENCES

- [1] Beaudoin CE, Tao CC. Benefiting from social capital in online support groups: an empirical study of cancer patients. *Cyber Psychol Behav* 2007;10(4):587–90.
- [2] Rodgers S, Chen Q. Internet community group participation: Psychosocial benefits for women with breast cancer. *Journal of Computer-Mediated Communication* 2005;10(4):1–27.
- [3] Wen KY, McTavish F, Kreps G, Wise M, Gustafson D. From diagnosis to death: a case study of coping with breast cancer as seen through online discussion group messages. *J Comput Mediat Commun* 2011;16(2): 331–61.
- [4] Balatsoukas P, Kennedy CM, Buchan I, Powell J, Ainsworth J. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. *J Med Internet Res* 2015;17(6):e141.
- [5] Masic I, Sivic S, Toromanovic S, Borojevic T, Pandza H. Social networks in improvement of health care. *Mater Sociomed* 2012;24(1):48–53.
- [6] Berkman LF. Social support, social networks, social cohesion and health. *Soc Work Health Care* 2000;31(2):3–14.
- [7] Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf* 2014;37(10):777–90.
- [8] Bakken S, Hyun S, Friedman C, Johnson SB. ISO reference terminology models for nursing: applicability for natural language processing of nursing narratives. *Int J Med Inform* 2005;74(7):615–22.
- [9] Feng D, Burns G, Hovy E. Extracting data records from unstructured biomedical full text. In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, June 28–30, 2007, Prague, Czech Republic, 2007, pp. 837–46.
- [10] McEwan R, Melton GB, Knoll BC, Wang Y, Hultman G, Dale JL, et al. NLP-PIER: a scalable natural language processing, indexing, and searching architecture for clinical notes. *AMIA Jt Summits Transl Sci Proc* 2016;2016:150–9.
- [11] Shashirekha HL, Murali S. Ontology based structured representation for domain specific unstructured documents. In: *Proceedings of the Conference on Computational Intelligence and Multimedia Applications* 2007; pp. 50–4.
- [12] Wen M, Rosé CP. Understanding participant behaviour trajectories in online health support groups using automatic extraction methods. In: *Proceedings of the 17th ACM international conference on supporting group work* 2012, Oct 27; pp. 179–88.
- [13] Denecke K, Bamidis P, Bond C, Gabarron E, Househ M, Lau AY, et al. Ethical issues of social media usage in healthcare. *Year Med Inform* 2015;10(1):137–47.
- [14] Mostafa MM. More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications* 2013;40(10):4241–51.
- [15] Mirhaji P, Byrne S, Kunapareddy N, Casscells SW. Semantic approach for text understanding of chief complaints data. In: *AMIA Annual Symposium Proceedings* 2006; p. 1033.
- [16] Wei T, Lu Y, Chang H, Zhou Q, Bao X. A semantic approach for text clustering using WordNet and lexical chains. *Expert Systems with Applications* 2015;42(4):2264–75.
- [17] American Nurses Association. ANA Recognized terminologies that support nursing practice; 2012. [online] [cit. 2017-11-01]. Available from: <http://www.nursingworld.org/MainMenuCategories/Tools/Recognized-Nursing-Practice-Terminologies.pdf>
- [18] Martin KS. The Omaha System: a key to practice, documentation, and information management (reprinted 2nd edition). Omaha, NE: Health Connections Press; 2005.
- [19] Kreitzer MJ, Koithan M (Eds). *Integrative Nursing*. New York: Oxford University Press; 2014.
- [20] Monsen KA, Peters J, Schlesner S, Vanderboom CE, Holland DE. The gap in Big Data: getting to wellbeing, strengths, and a whole-person perspective. *Glob Adv Health Med* 2015;4(3):31–9.
- [21] Bird S, Klein E, Loper E. *Natural language processing with Python: analyzing text with the natural language toolkit*. Boston: O'Reilly Media, Inc.; 2009.
- [22] Olsen AM, Baisch MJ, Monsen KA. Interpretation of ecological theory for physical activity with the Omaha System. *Public Health Nurs* 2017;34(1):59–68.
- [23] Cueller NG. Integrative nursing management of sleep. In: Kreitzer MJ, Koithan M (Eds). *Integrative Nursing*. New York: Oxford University Press; 2014, pp. 230–42.
- [24] Wickwire EM. Financial costs of insomnia. *Sleep Review: Journal for Sleep Specialists*; 2014. [online] [cit. 2017-11-01]. Available from: <http://www.sleepreviewmag.com/2014/12/financial-costs-insomnia/>