Wilson, Max L. and Ali, Susan Abubakir and Valstar, Michel F. (2014) Finding information about mental health in microblogging platforms: a Case study of depression. In: 5th Information Interaction in Context Symposium (IIiX'14), 26-29 August 2014, Regensburg, Germany.

Finding Information about Mental Health in Microblogging Platforms: a Case Study of Depression

Max L. Wilson, Susan Ali, Michel F. Valstar
Mixed Reality Lab
School of Computer Science
University of Nottingham, UK
{max.wilson, itxsaali1, michel.valstar}@nottingham.ac.uk

ABSTRACT

Searching for online health information has been well studied in web search, but social media, such as public microblogging services, are well known for different types of tacit information: personal experience and shared information. Finding useful information in public microblogging platforms is an on-going hard problem and so to begin to develop a better model of what health information can be found, Twitter posts using the word "depression" were examined as a case study of a search for a prevalent mental health issue. 13,279 public tweets were analysed using a mixed methods approach and compared to a general sample of tweets. First, a linguistic analysis suggested that tweets mentioning depression were typically anxious but not angry, and were less likely to be in the first person, indicating that most were not from individuals discussing their own depression. Second, to understand what types of tweets can be found, an inductive thematic analysis revealed three major themes: 1) disseminating information or link of information, 2) self-disclosing, and 3) the sharing of overall opinion; each had significantly different linguistic patterns. We conclude with a discussion of how different types of posts about mental health may be retrieved from public social media like Twitter.

Categories and Subject Descriptors

H.3.4 [**Information Systems**]: Information Storage and Retrieval—*Systems and Software*; H.5.m [**Information Systems**]: Information Interfaces and Presentation—*Misc.*

Keywords

Microblogging, Information Seeking, Information Retrieval, Mental Health, Depression

1. INTRODUCTION

Search for online health information [41], and search within social media sites, including microblog search [11], are two on-going hard problems. Yet while more is known about how people search for health-related information online, we know much less about health information that can be found in public microblogging services like twitter [10]. There are two reasons that make these public social media platforms a valuable source of online health information. First, social media provide an important medium for people with health concerns to communicate with other sufferers [24]. Second, microblog services are well known as a source of tacit information that is less common online: personal experience [19]. Consequently, this research focuses on understanding the different types of posts that can be found in microblog services about health-related issues.

Although De Choudhury et al investigated which health conditions are discussed in Twitter [10], to study the *types* of posts and information that might be retrieved about different health concerns from microblogging services, we have focused on a case study of a prevalent mental health issue, depression. Depression has increased significantly both in developed and developing countries [2] over recent years and is estimated to affect over 350 million people [42]. Discussing the problems related to depression with others is accepted to be an important element of coping, but personal factors often discourage people from doing so face-to-face [1]. Social media, therefore, provides an convenient platform for people to communicate and interact with potentially millions of people from different countries and backgrounds [3], whilst reducing the negative connotations with face-to-face disclosure [12, 28]. Consequently, we see the analysis of tweets about depression, rather than by people with depression [32], as a proxy for studying what forms of information may be retrieved about health issues from microblogging services.

Existing research has investigated the use of social media by those with mental health disorders such as insomnia [20] and autism [17]. De Choudhury et al [8] have shown that tweets can be used to objectively but unobtrusively identify postpartum depression, and also predict when users are entering a period of depression [9]. Further, Park et al. [32] compared the behaviour of people with and without depression on Twitter. Although these studies have investigated how these health issues manifest on microblog services, they have not studied the broader range of information about the health issues that can be found. We contribute: 1) a linguistic, and 2) a thematic analysis of information about mental health that can be found on public microblog services, using a case study of posts about depression on Twitter.

2. RELATED WORK

Many people search for health information online. Accord-

IIiX '14 August 26 - 29 2014, Regensburg, Germany

http://dx.doi.org/10.1145/2637002.2637006.

ing to Fox and Jones [13], using online information about health can have a potential positive effect on the user's health. Recent research has focused on the importnace of health information retrieval, with Morris and Morris [27] suggesting that researching health problems may better support patients and doctors diagnose problems collaboratively. White and Horvitz [41] found, however, that people are more likely to develop notable anxiety, which they call Cyberchondria, than find useful and accurate information about their health problems. Consequently, the accuracy and authority of health information online has become a key source of relevance for health search. [43].

Many people also search for health information from social media sites. Approximately 70% of Canadian adults, for example, use social media to search for information, particularly health information. From that 70%, it was found that women have more interest to search for information about health and medical information (74%) compared to 66% of men [38]. De Choudhury et al compared the information that people search for on Bing and Twitter for a large range of health conditions, and found that Twitter was a commonly used platform for low-stigma problems and symptom oriented information [10]. Consequently, this work has focused on investigating the types of information that can be found about health problems in social media, using depression as a case study, due to its high prevalence.

2.1 Major Depressive Disorder (MDD)

MDD is one of the most serious and prevalent types of mental health disorders that can affect individuals at any stage of their life [2]. From the perspective of Psychology, there are many factors that might lead to depression such as lack of face-to-face social communication, problems in personal life, loneliness, and loss of job. Moreover, depression can cause further mental problems ranging from sadness, anxiousness, low self-esteem, etc. Therefore, these moods and feelings might lead to other symptoms such as inability to work, eating disorders, lack of ability to concentrate, and more seriously, in some cases, depression might lead people to disabilities or suicide [42].

Depression has increased significantly over the last 20 years in both developed and developing countries. An estimated 16% of the population suffer from depression; it is considered to affect 1 in 10 adults one or more times in their entire life [23]. Researchers have identified that the population in western and wealthy countries are affected by depression more significantly than those in low-income countries in the Middle East and Asia [31]. Depression can be effectively treated in a number of ways including pharmacological treatment, psychoeducation, (e.g. improving lifestyle) and cognitive behavioural therapy. However, Lepine and Briley [24] argued that people who have depression should communicate, disclose and discuss their problems with others to obtain support from them. Unfortunately, many people find it difficult to disclose their depression face-to-face due to social or cultural factors [28]. Therefore, people who suffer from mental health disorders may suffer from "information poverty" [16] and lack of support. For these reasons, people with depression often cope with their mental health concerns by asking for help or disclosing their problems with people they *do not know*, or who have similar experiences [30].

2.2 Mental Health and Social Media

The study of mental health issues manifesting in social media is not uncommon. Moreno et al [26], for example, studied how depression manifests in Facebook. According to Weinberg [40], however, networks like Facebook involve more personal commitments and social relationship to maintain friendships, find and contact old friends, or share videos and photos with them. In contrast, public networks like Twitter allow users to write short messages on different topics broadly to any user even if they are not friends. Moreover, Twitter allows users to ask questions and receive a response quickly and to obtain knowledge about the latest news and events [12]. Morris et al. [28] found that the most frequent type of questions asked on social media were recommendations followed by opinions. Question's topics primarily included: entertainment, personal, and health questions, but personal and health questions received fewer responses [33]. On Twitter, users are allowed to follow anybody without asking permission and tweets are typically public, making it a potential source of valuable experience-oriented information [19]. Moreover, hashtags and retweets allow users to find a similar topic and distribute tweets of other users easily [4, 37], but there are many challenges associated with microblog search that make finding such information, outside of your own personal twitter network, difficult [11].

Hasler and Ruthven [16] discovered that people with sensitive health problems tend to ask help from people who have experienced a similar situation rather than close friends and family. Further, they found that people can find it difficult to disclose health concerns on public social media, and so preferred anonymous forums that allowed the freedom to safely ask questions. For the same reasons, many Twitter users do not use real names for their account in order to have freedom for discussing more sensitive topics [18]. Jamison-Powell et al. [20] found that people often discuss their insomnia on twitter, share their experiences, and distribute information. They found that tweets containing "insomnia" included more negative health information and made significant reference to "time", and particularly "present tense" experiences.

Focusing on MDD itself, Park et al. [32] interviewed people with depression, and compared their twitter usage to people without depression. They found that people with depression tended to follow users who tweeted about their daily lives rather than following someone who tweeted about depression and gloom. This finding is similar to Kuiper and MacDonald's [21] study, as they found that individuals with depression became more depressed after they had contact with other depressed patients (by phone). Conversely, Park et al. found that all participants with depression preferred Twitter to Facebook because of the loose social connection that allowed them to tweet more openly. In studying postpartum depression, De Choudhury et al. [8] retrieved tweets of new mothers and showed that it was possible to determine notable sentiment shifts either-side of a tweet announcing a birth, using features from a linguistic analysis, such as positive and negative emotional terms. Further, De Choudhury et al [9] discovered it was possible to predict episodes of depression, using features such as change in posting frequency and increased concern over health issues. Each of these studies have investigated people with mental health concerns, rather than what forms of information, tacit or otherwise, can be found online about mental health.

Table 1: Linguistic Comparison between Depression and General Sets of Tweets

Word Class	Depression Tweets[a]		Non-specific Tweets[b]		Independent t-test	
	Mean	Std	Mean	Std	p	t
Character Count	15.2704	6.69879	10.7058	6.67133	P<0.001	57.064
Pronoun	8.4709	8.68415	11.4075	11.41211	P<0.001	-24.369
I	4.4635	6.53561	5.9077	8.49577	P<0.001	-16.029
Positive Emotion	2.2934	5.18795	3.5381	7.74198	P<0.001	-15.942
Negative Emotion	11.3374	10.69361	2.6860	7.30208	P<0.001	78.222
Anxious	0.9441	3.64111	0.1437	1.61091	P<0.001	23.355
Anger	1.1728	4.86862	1.3870	5.32722	P<0.001	-3.499
Sad	8.5309	8.13062	0.4468	2.83045	P<0.001	108.785
Time	3.2745	5.75745	3.8717	8.38312	P<0.001	-7.005
Past	1.4923	3.64448	1.9080	4.91159	P<0.001	-7.972
Present	9.9159	9.30484	10.5457	11.09264	P<0.001	-5.118

[a]n=13279, [b]n=14727, Std= Standard Deviation

3. ANALYSIS OF DEPRESSION TWEETS

To understand the nature of available information on Twitter, we decided to collect a corpus of tweets to analyse both linguistically for features that make them recognisable, and then manually to discover themes.

3.1 Data Collection

Two corpora of 25,000 tweets were gathered between the 28th June and 2nd July 2013. The first corpus was collected by searching for the keywords "depression" or "dep". These terms represent a first-step query for what people are saying about depression on twitter; later, we discuss the implications of focusing on just these two search terms. The second corpus was an unbiased sample from the open Twitter stream. To remove duplicate content, straight retweets (without additional content from the retweeter) were removed. Since this study focuses on the tweets, posts that were retrieved because they included "depression" or "dep" in the author's username were also deleted. 13,279 tweets remained in the depression group and 18,280 tweets in the non-specific group. To make the groups more comparable, a stratified sample of tweets was removed from the non-specific group, leaving 14,727 tweets. To preserve anonymity, all twitter handles were replaced with "...", so we can see what was said, but not to or about whom.

To explore the language being used in the two corpora, to see whether a corpus of tweets about depression was inherently different from a general sample, Linguistic Inquiry and Word Count (LIWC) software [6] was used to analyse the content. This software identifies certain linguistic patterns of the tweets, categorises them in a psychologically meaningful way according to different word class and provides the percentages of word class [34]. Of the many categories, the most notable differences were found in use of pronouns, forms of emotion, and in the tense of the language. These are prioritised in our analysis below, but a full analysis of the depression set is shown in Appendix A.

3.2 Linguistic Analysis

Table 1 shows the key features of the two groups; the means are the average distribution of that category in the tweets, where each tweet was analysed separately. First, the depression group included significantly more characters than the tweets of the non-specific tweets, indicating that their length is a possible factor of relevance for retrieval. It was found that there was a significant difference of using pronoun and "I" between the two groups; the non-specific tweets contained more pronouns and "I" than the depression tweets ($p<0.001$). The depression group also referred to time significantly less than the general tweets ($p<0.001$). Both past and present tenses were mentioned less frequently in the depression tweets.

The analysis found a significant difference between the two groups in terms of sentiment. Perhaps predictably, negative emotions were used significantly more in the depression tweets ($p<0.001$) compared to the positive emotions when found in the non-specific tweets. The sadness category, followed by the anxiety category, were the most common amongst the negative content. They were also used significantly more when compared with the non-specific tweets ($p<0.001$). Conversely, the anger emotion was found more frequently in the non-specific group. For investigating the relationship between these sentiments, a Pearson correlation showed that there was a significant weak-positive relationship between the sadness and anxiety groups ($r=.041$, $p<0.001$) and a weak-positive relationship between the sadness and anger groups ($r=.036$, $p<0.001$).

To summarise, the depression tweets contained more characters and more negative emotions (sadness and anxiety) compared to the non-specific tweets. Further, the depression tweets contained less positive emotion, fewer pronoun and "I" occurrences, less anger, and were less pre-occupied with time, including both past and present tenses. The finding that the depression tweets contain fewer pronouns and use of "I" is notable, given that prior work has found people with depression use the term "I" more often. This indicates that tweets using the word "depression" are not necessarily produced by those with depression. To investigate this finding, and further learn about the different types of information that can be found in microblog services like Twitter, we performed a thematic analysis using Grounded Theory, described below.

4. THEMATIC ANALYSIS

Although the linguistic features help us to recognise that the majority of depression tweets are time-independent and negative but not angry, we still know very little what they were actually about. For example, tweets such as: *"Again lovely depression here"* is coded by LIWC as involving positive emotion, whereas it is actually negative. To investigate the types of tweets about depression, we performed a manual qualitative Grounded Theory analysis [15] of a sample from the depression group. We adopted the inductive Glaserian approach [14] to generate a comprehensive model from the data, rather than trying to validate or develop a pre-existing taxonomy. This allows us to start without assumptions about the types of tweets relating to mental health issues like depression, and focus on what the data tells us.

4.1 Methodological Process

We began by selecting an initial sample of 250 tweets from the depression group. The sample was read several times by two researchers, independently, and multiple approaches to selecting the focus of coding were considered. After choosing an approach, one researcher applied open coding to the 250 tweets and produced 84 codes. A selection of these codes was then sampled by a second researcher, and axial and selective coding was applied to further review the codes. This process led to agreement between the researchers, and we began the process of merging the codes into categories and themes. After several iterations of refinement, 27 key categories were identified within 8 first-order themes, and in 3 second-order themes.

In order to validate the themes and categories produced through our refinements, a new sample of 25 tweets were collected per day from the 5-day depression corpus. These additional 125 tweets were first analysed by the first researcher. No new or differentiating codes were produced, and so the same sample was provided to an independent researcher. The independent researcher was also given a copy of the identified first- and second-order themes, where each had a definition and an indicative tweet. The judgements from the primary and independent researchers were compared using Cohen's Kappa analyses. A kappa score of 0.772 was achieved for the first-order themes, and a kappa of 0.813 was reached for the second-order themes. These scores are considered "substantial agreement" and "almost-perfect agreement", respectively [22]. We took this to indicate that the taxonomy was stable and could be applied consistently. Finally, to linguistically analyse a more comprehensive sample, the primary researcher manually classified a further 1,000 tweets. In total from all three phases, and after removing junk content, 1,130 depression tweets were classified into one of the 8 first-order themes.

5. RESULTS

Table 2 shows our final taxonomy, broken into 8 first-order themes and 3 second-order themes. Each first-order theme has an example, and a list of the most common sub-categories found in that set. The three larger second-order themes, which had critical mass for statistical comparison, were analysed for the same 10 linguistic features analysed in Table 1. An ANOVA was used to compare the sets for each linguistic feature. Each theme is described in more detail below, and the linguistic features of the second-order themes are then discussed.

5.1 Disseminate Information or Shared Link

This second-order theme represents tweets that were posted for the purpose of disseminating information either through their textual content or by providing a link.

5.1.1 Depression Fact

Depression Fact illustrates the way that users provided information about what depression is, the impact of depression, and the risk of depression on a person's life. These tweets often included statistic figures or ratios to provided more accurate and real information, e.g. *"80% of depression and stress are work related, while 99% of labor is cursed: Taxes and tithes are redemptive entities."*. Although cynical in nature, an example that tried to make more people aware about depression: *"why does every1 beg depression LOL it's a mental illness not a fashion accessory get out man"*. Some tweets were about the impact of depression: *"People think depression is just being sad but it ruins every part of your life."*

5.1.2 Cause of Depression

The **Cause of Depression** first-order theme included tweets that try to raise awareness about the potential factors that may cause depression. For instance, lack of sleep was the most common cause mentioned in tweets: *"Not sleeping enough leads to desire for sex, depression and alcoholism"*. In addition, users tweeted about the risk of experiencing negative feelings and its consequent impact *"Being bored so much, it turns into depression"*. Moreover, some provided information about external factors that lead to depression such as workload: *"Too much homework can cause stress, depression, and even lower grades."*

5.1.3 Depression Factors and Managing Depression

Depression Factors and Managing Depression included tweets posted by the users with the aim to support people and how to manage depression. For this reason, users have attempted to provide information about several factors that relate to depression such as human or dietary factors: "#Depression and aging: A growing trend http://t.co/yYayx4ThNo" and *"Study Confirms Fast Food and #Depression Link http://t.co/mptcgfjXOx."*. In terms of managing depression, tweets described a wide range of approaches and strategies that can be used to manage depression. Natural solution was one of the most frequent approaches for managing depression that reported by the users: *"@...socialising, exercising and sunlight are good treatments for depression."*. Further, some users suggested using online services to manage depression: *"Online Counseling Service available via Skype-Helping you manage #ANXIETY & #DEPRESSION. http://t.co/z6OtceRBgi. Email me. Please Retweet!"*. Some users suggested links to websites for identifying how depression can be diagnosed, tested and managed. The difference between this category (supporting people) and online service is that users in the online ser-

Table 2: Taxonomy of Tweet Types in the Depression Set, with Linguistic Comparison of 2nd-Order Themes

Reason of Tweet (2nd-order Theme)	Content of Tweet (1st-order Theme)	Categories, Examples, and Ling. Features (Average % of Content)									
		Pronoun	I	Positive	Negative	Anxious	Anger	Sad	Time	Past	Present
1) Disseminate Information or Information Link	1.1) Depression Fact	**Categories include:** Ratios, Facts, and Related Issues.									
		Example: "At least 1 in 4 people have or will struggle with depression in their lifetime ... #mentalillness... http://t.co/Zpu9FDHDTy"									
	1.2) Cause of Depression	**Categories include:** External Factors, Negative Feelings, and Sleep issues									
		Example: "Too much homework can cause stress, depression, and even lower grades." **And:** "Not sleeping enough leads to desire for sex, depression and alcoholism"									
	1.3) Factors and Management	**Categories include:** Human Factors, Diet, Natural Solutions, Online Service, and Support People									
		Example: "@...socialising, exercising and sunlight are good treatments for depression." **And:** "Online Counseling Service available via Skype - Helping you manage #ANXIETY & #DEPRESSION. http://t.co/z6OtceRBgi. Email me. Please Retweet!"									
	Ling. Features	5.75%	0.8%	5.16%	11.4%	2.71%	0.6%	7.6%	2.7%	0.4%	4.83%
2) Self Disclosing	2.1) Coping with Depression	**Categories include:** Attempts to Avoid, and Eduring Depression									
		Example: "@...YEY! Dont worry,Im too excited! Is it on E4? I'll put on +1! Gather il feel a slight depression after too #doesnthappeninreallife"									
	2.2) Having Depression	**Categories include:** Describing the Experience, Feelings, Environment, Unknown Reasons, and People									
		Example: "Stress + depression = worse thing ever - I just wanna cry.." **And:** "feeling like i've gone into some depression mode for no damn reason" **And:** "My biggest depression rn is my little sister"									
	2.3) Opinion	**Categories include:** Expecting Depression, People, and Social Factors									
		Example: "If this weather keeps up all of windsor may go into a state of depression" **And:** "a lot of people ask why im so open about the fact i have depression and im like why not.. i dont blame myself for it."									
	Ling. Features	13.6%	4.02%	4.61%	13.2%	0.69%	2.1%	9.27%	5.2%	1.7%	12.2%
3) Social Engagement	3.1) Discussing Depression	**Categories include:** Environment, Feelings, Opinion of People, General Opinion, Attitude, and Sources									
		Example: "Are the holidays going to cheer you up or make your depression worse?" **And:** "Anyone who has depression is automatically beautiful to me" **And:** "People miss guide my thoughts for depression its not my thoughts, they've never changed only my emotions, these pills make me feel great tho"									
	3.2) Sharing Support	**Categories include:** Natural Solution, Available to Help, and Doing Activities									
		Example: "To everyone out there struggling with something, wether its depression, self harm, an eating disorder, whatever it is. I am here for you."									
	Ling. Features	16.7%	11.9%	2.46%	12.5%	0.7%	1.21%	9.28%	6.67%	2.1%	11.8%
	F-Scores	192.27	518.42	22.66	3.79	42.42	14.34	8.97	43.85	27.52	126.12
	P-Values	<0.001	<0.001	<0.001	<0.05	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

vice category posted information about managing depression along with suggesting online services such as Skype instead of providing a link to websites where people can find useful information.

5.2 Self Disclosing

This second-order theme was formed from tweets by users who were disclosing information about and experiences from their own depression. Users had different reasons for disclosing their depression on Twitter.

5.2.1 Coping with Depression

Coping with Depression referred to tweets that were posted by users who disclosed information about how they cope with depression. For example, some of the users tried to avoid being depressed: *"I need to leave this house again. It's alot of depression in this house. I don't wanna be around That mess."* In addition, a few users tweeted about using Television or medical treatment for coping with their depression: *"I need to go to the doctors and get some depression tablets"*.

5.2.2 Having Depression

Within **Having Depression** it was found that many users disclosed their depressed condition, the factors that led to their depression and how they felt when depressed. Some users tweeted only to disclose or announce their depression: *"This sadness and depression are a part of me"*. Several users disclosed their feelings of being depressed due to people for instance: *"My biggest depression rn is my little sister"*. As it can be seen, many assigned the cause of their depression to certain people. Some users disclosed their feelings of depression without having any specific reason: "feeling like I've gone into some depression mode for no damn reason". In addition, some users described their personal feelings towards depression: *"Stress+depression=worse thing ever - I just wanna cry."* Others described their experience with depression: *"Still suffering from post concert depression"*. Further, several users disclosed their depression and related it to environmental factors: *"@...:(I'm falling into a deep depression. This weather needs to change asap!"*

5.2.3 Opinion

Opinion represents the tweets of users who disclosed their depression by providing their own opinion of it. For instance, some users disclosed their opinions about expecting depression: *"If this weather keeps up all of [winter] I may go into a state of depression"*. In addition to self-disclosing, users in this theme have also disclosed their opinion about other people's opinions: *"a lot of people ask why im so open about the fact i have depression and im like why not..i dont blame myself for it"*. Also, some users tweeted their opinion about social factors (people who related to their depression either positively or negatively): *"@...Thank you for that tweet, Wil. Depression has me your inspiring tweet gives me hope that it won't someday"*.

5.3 Social Engagement

This theme included tweets that were posted for providing opinions about depression in general. In this theme, users did not disclose or refer to having depression, unlike the "opinion" sub-theme in the theme of "self-disclosing".

5.3.1 *Discussing Depression*

In **Discussing Depression**, users wanted to give opinions about depression and to discuss issues related to depression. From the tweets, it was found that the majority of users reported their opinion about people. For instance: *"Anyone who has depression is automatically beautiful to me"*. There was also opinion about the behaviour of people: *"People who complain online are more likely to suffer from anxiety depression and stress. 'Okay' "*. In addition, some users posted their opinion about other people's opinion and attitude towards depression: *"hate when folk throw the word depression around like its nothing"*. Some users reported their thoughts and feelings about depression in general without referring depression for themselves. In addition, some users provided their opinion about the source of the depression *"@...yeaa problems do cause depression...cheer up"*. Tweets were often about different attitudes towards depression. For instance, they described their attitude while reading about depression: *"I would drink while reading my book about anxiety and depression - bahaha #fuckmylife."*

5.3.2 *Sharing Support*

Sharing Support illustrates how users commented on the strategies and approaches for managing depression. Although the passing on of recognised information was noted in the "disseminate information" theme, this theme relates to people's own advice and experiences of whether strategies were successful or not. From the tweets, it was found that most users in this theme commented on natural solutions for managing depression: *"@...Studies show that sex helps you live longer, makes you smarter and prevents depression. I agree with that.:)"*. In addition, some users voluntarily offered to support people who suffer from depression: *"To everyone out there struggling with something, wether its depression, self harm, an eating disorder, whatever it is. I am here for you"*. A common occurrence was that tweets in this category required some form of self-disclosure in order to provide advice, such as: *"I think Best stuff against depression, lonelyness, despair and sadness: Drinking Alcohol #GreatStuff"*. Advice in this category, however, was not always in agreement with professional advice.

5.4 Linguistic Differences between Themes

Even with only 1130 tweets analysed linguistically, the results indicate some significant differences between the content in each theme. These linguistic differences are shown as part of Table 2. Negative emotion was the most consistent linguistic feature across all themes. Sadness was also fairly consistent across the second-order themes, but less present in the Dissemination tweets. First person language, however, and particularly the use of I, were much more prevalent in the Social Engagement tweets, and least present in the Dissemination tweets. This means that use of the first person could be used to differentiate between types of depression tweet. To detect examples of self-disclosure, though, present tense was a much clearer indicator.

6. DISCUSSION

This research has used two approaches to produce a rounded picture of the types of depression information that are available on twitter. Below we discuss how these findings help inform the types of posts, or information, that might be retrieved when users searcher for mental health information in microblogging platforms.

6.1 Thematic Findings

The thematic analysis found three different key types of tweets, which can be broken down into 8 smaller groups. Consequently, when searching for information about depression, searchers may encounter each of these types without comprehending their different values. Some types were more informational, as per the disseminate theme, but the self-disclosing themes involved users talking about potentially real and relevant experiences. Providing searchers with information about these different major types of tweets may help them to either be more prescriptive of what they are looking for, or be more aware of what value each can provide.

From the results of the Disseminate theme, it can be argued that some tweets are from users who voluntarily disseminate or pass on valuable information, about its cause and effects and factors that directly or indirectly relate to depression. Further, they tended to raise the awareness of people about the approaches that can be used to manage depression and to minimise its potential risks. Twitter users have not only supported others by providing information in their tweets, but they have also offered support through suggesting links to websites that might be useful for those who tend to know more about depression. The most common information that has been disseminated was related to managing depression through a natural solution such as exercise followed by providing information about online services. One concern, especially for the dissemination of links, is that there is a chance that tweets are acting as a form of spam linking to websites and blogs.

These results around dissemination, however, mirror some findings from other studies in which it has been noted that useful information or links to other information have been widely posted by many users [4]. In addition, Naaman et al. [29] have described Twitter as a "social awareness stream" since it has been used widely by people who tend to make each other aware about what they know regarding a wide range of topics. As a result, it is worth mentioning that users might be engaged in this information dissemination due to the "public" feature of tweets. Since a large number of public tweets are posted about depression, this may have persuaded people who suffer from depression to use Twitter. These results also support findings from Scanfield et al. [37] who argue that the public feature of Twitter has led users to disseminate information and provide support to a public audience. Consequently, people will become increasingly likely to use such platforms to find potentially useful information.

From the Self Disclosing theme, it would seem that users who suffered from depression do sometimes use Twitter as a platform to disclose their depression, how to cope with it, and their thoughts about depression. This attitude of self-disclosure might be due to the fact that there is an element of real-time on Twitter, which allows people to write about their condition, feelings and thoughts at the time when they use Twitter; self-disclosing tweets were more frequently in the present tense. In addition, the "public" feature facilitates sending tweets to unknown audiences without specifying any particular tweeter and without the requirement of making friendships. According to researchers, communicating with people who have had similar experiences, and sharing information with them is considered a common strength in online support-groups [7]. Consequently, users who suffer from depression might become able to communicate with similar users and act as an "adviser", with lower perceived risks than face-to-face discussions [25]. In this way, the large global nature of Twitter facilitates discussion or disclosure of sensitive topics like depression. Further, loose restrictions on anonymous accounts allow people to achieve anonymity within this larger audience. Hasler and Ruthven [16] found that anonymity in dedicated health forums was a significant factor in their self-disclosure. This category of tweets, therefore, may be valuable for hearing about people with similar experiences, or for finding them to create or integrate within a loosely-tied support network.

The key findings in the Social Engagement theme show that the users tweeted to provide their own opinions and thoughts about depression, rather than trying to disseminate formal or official information. In addition, users have provided their opinions and comments about the approaches that can be used to manage depression without mentioning whether they have depression or not. Although some opinions may be more or less useful than others, some appeared to be based on experience, which Hurlock and Wilson found to be a common form of useful tweet. Moreover, some users offered their personal availability to support people who suffer from depression instead of suggesting advice and information. These results indicate that forums like Twitter allow people to more freely share information, experience, and opinion in a more integrated form, and perhaps with equal value. Although there have been many useful comments about depression, public forums like Twitter also allow people to share inappropriate or incorrect advice. For example, Sullivan et al. [39] found that incorrect information and advice was disseminated (via Twitter) to people with sports related concussion. The correctness of information in social media is a significant challenge in research on its own [35].

6.2 Limitations and Beyond this Case Study

Although our studies have produced novel insights into forms of information about depression on Twitter, more work could be done to relate these findings to those in other forms of social media. Certain results may remain the same, while elements such as the public nature of Twitter, or the loose-ties created by following, that are unique. Although we analysed a large number of tweets, these were primarily retrieved using the "depression" and "dep" keywords, where "dep" was noted as a common keyword used in tweets about depression. We acknowledge that depression may be discussed in tweets without using these specific keywords; Park et al [32] investigated the tweets of interviewees with depression. We take these two terms, however, as the most likely first query used to search for posts about depression on Twitter. Further, we acknowledge that lots of tweets may not be really about MDD, but using the word depression informally. This further highlights, however, the challenge for microblog health search invoked by using these search terms.

One important route for further work is to understand which issues are specific to depression and which help us understand microblog search for mental health information in general. One approach is to corroborate our findings with similar studies of other mental health issues in tweets (e.g. tweets about insomnia [20]). Additional case studies, and corroborating findings will help us to build a general model. We discuss some of the differences between our findings, and the findings of other work for microblog mental health information retrieval below.

6.3 Implications for Retrieval Systems

Hurlock and Wilson [19] discuss the difference between tweets being relevant and being useful, and have identified features of tweets that distinguish the more and less useful. The themes identified above, combined with the linguistic analysis, may support such efforts to improve searching with social networks by helping to identify useful tweets about depression. One result of our linguistic analysis was that the majority of tweets were depersonalised away from pronouns and the use of "I". These findings differ from previous studies in that they found people who suffer from depression use more first-person pronouns in other media than people who do not have depression [36, 6]. Burke et al. [5], for example, stated that the probability of getting answers to questions could be increased by using "I" during self-disclosure.

More specifically, depression tweets contained more sad and anxious words. However, anger was mentioned more commonly in the non-specific tweets. For a system trying to retrieve tweets about mental illness, a language model could be developed that focuses on these differences, in order to improve the accuracy of the results. Our findings, however, are different to the manifestation of some other mental health problems, where Jamison-Powell et al. [20] found that anger was mentioned more commonly in tweets about insomnia. Jamison-Powell et al. found other differences, like that insomnia was described frequently in relation to time, but depression tweets used significantly less language about time than the general group. This perhaps reflects that depression affects sufferers more consistently, at all times, while episodes of insomnia may be extended but affects people when trying to sleep. These differences indicate that, where possible, a language model could be optimised towards a type of mental illness.

We also saw some variation between the groups discovered in the thematic analysis, which would help to find more specific forms of tweets about depression. Dissemination tweets are less frequently in the first person, and less concerned with time. Dissemination tweets had notably fewer pronouns in them, which may be a clear indicator. Self-disclosing tweets were more frequently in the present, were more angry, and used more pronouns (but less use of "I") than other depression tweets. Finally, Opinion tweets, while also having more pronouns, used the word "I" more frequently, and included more sadness-related and past-tense language. Knowing the biases of certain language features, towards types of tweets, means that a more complex IR language model could be developed. This would allow, for example, for a system to find more examples of actual experience of depression, by boosting the weight of terms relating to present tense, anger, and use of pronouns. Conversely, the language model could better find information-focused tweets, by bossting the weight of positive and anxious terminology. Whilst past-tense personal pronouns use could be boosted in the language model for social discussions of depression.

7. CONCLUSIONS

This paper has presented a mixed-method combination of two analyses aimed at understanding what types of information are available on public social networks like twitter. A broad linguistic analysis revealed that tweets about depression can be identified as being significantly less in the first person, more negative but less angry, and less preoccupied by time, than a general set of tweets. These findings indicate that many were perhaps predicatably negative, but not all were oriented around self-disclosure. Second, a thematic analysis identified 8 first-order themes that describe distinct types of tweets according to 3 main intentions behind tweets posted on twitter about depression: sharing information, self-disclosing, and providing opinion. Each of these had different linguistic features, where self-disclosing tweets, for example, were more in the first person, as found in prior work focused on social media users with depression. Together these two sets of results provide both information about the types of tweets about depression that can be found, and their individual identifying features to recognise them.

Our research has provided 3 main contributions: 1) the linguistic analysis of a corpus of depression tweets in comparison to a general sample, 2) a taxonomy of types of tweets that are available online, with descriptions and examples, and 3) recommendations for retrieving different types of information about Depression, as a first step towards generalising to general health information retrieval in microblog search. In our own research, we aim to support people in both sharing and finding real experiences of how they manage to implement suggested treatments, like talking more about their depression, or avoiding being lonely. Such practical advice is often missing from formal but general guidelines.

8. ACKNOWLEDGMENTS

We would like to thank the Kurdistan Regional Government (Ministry of Higher Education and Scientific Research) for funding the MSc studentship that supported this project. The work of Max Wilson and Michel Valstar is partly funded by Horizon Digital Economy Research, RCUK grant EP/G065802/1. In addition, Michel Valstar's work is partly funded by the NIHR-HTC 'MindTech'.

9. REFERENCES

[1] M. D. Back, J. M. Stopfer, S. Vazire, S. Gaddis, S. C. Schmukle, B. Egloff, and S. D. Gosling. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21(3):372–374, 2010.

[2] BBC-Health. '1 in 5' children suffer from depression, 2013.

[3] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[4] d. boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, HICSS '10, pages 1–10, Washington, DC, USA, 2010. IEEE Computer Society.

[5] M. Burke, E. Joyce, T. Kim, V. Anand, and R. Kraut. Introductions and requests: Rhetorical strategies that elicit response in online communities. In *Communities and Technologies 2007*, pages 21–39. Springer, 2007.

[6] C. Chung and J. W. Pennebaker. The psychological functions of function words. In K. Fiedler, editor, *Social communication*, pages 343–359. Psychology Press New York, 2007.

[7] N. S. Coulson, H. Buchanan, and A. Aubeeluck. Social support in cyberspace: a content analysis of communication within a huntington's disease online support group. *Patient Education and Counseling*, 68(2):173–178, 2007.

[8] M. De Choudhury, S. Counts, and E. Horvitz. Major life changes and behavioral markers in social media: case of childbirth. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW'13, pages 1431–1442. ACM, 2013.

[9] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, ICWSM'13. AAAI, 2013.

[10] M. De Choudhury, M. R. Morris, and R. W. White. Seeking and sharing health information online: Comparing search engines and social media. In *Proceedings of the 32nd Annual SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1365–1376, New York, NY, USA, 2014. ACM.

[11] M. Efron. Information search and retrieval in microblogs. *Journal of the American Society for Information Science and Technology*, 62(6):996–1008, 2011.

[12] M. Efron and M. Winget. Questions are content: a taxonomy of questions in a microblogging environment. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–10, 2010.

[13] S. Fox and S. Jones. *The social life of health information*. 2009.

[14] B. G. Glaser. *Theoretical sensitivity: Advances in the methodology of grounded theory*, volume 2. Sociology press Mill Valley, CA, 1978.

[15] B. G. Glaser and A. L. Strauss. *The discovery of grounded theory: Strategies for qualitative research*. Transaction Books, 2009.

[16] L. Hasler and I. Ruthven. Escaping information poverty through internet newsgroups. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, ICWSM'11, pages 153–160. AAAI, 2011.

[17] H. Hong, J. G. Kim, G. D. Abowd, and R. I. Arriaga. Designing a social network to support the independence of young adults with autism. In *Proceedings of the 2012 Conference on Computer Supported Cooperative Work*,

CSCW'12, pages 627–636. ACM, 2012.

[18] D. J. Hughes, M. Rowe, M. Batey, and A. Lee. A tale of two sites: Twitter vs. facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2):561–569, 2012.

[19] J. Hurlock and M. L. Wilson. Searching twitter: Separating the tweet from the chaff. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, ICWSM'11. AAAI, 2011.

[20] S. Jamison-Powell, C. Linehan, L. Daley, A. Garbett, and S. Lawson. I can't get no sleep: discussing# insomnia on twitter. In *Proceedings of the 30th annual SIGCHI conference on Human factors in computing systems*, CHI'12, pages 1501–1510. ACM, 2012.

[21] N. A. Kuiper and M. R. MacDonald. Self and other perception in mild depressives. *Social Cognition*, 1(3):223–239, 1982.

[22] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[23] S. Lee, J. Jeong, Y. Kwak, and S. K. Park. Depression research: where are we now? *Molecular brain*, 3(1):8, 2010.

[24] J.-P. Lepine and M. Briley. The increasing burden of depression. *Neuropsychiatric disease and treatment*, 7(Suppl 1):3, 2011.

[25] A. E. Marwick and d. boyd. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1):114–133, 2011.

[26] M. A. Moreno, L. A. Jelenchick, K. G. Egan, E. Cox, H. Young, K. E. Gannon, and T. Becker. Feeling bad on facebook: Depression disclosures by college students on a social networking site. *Depression and anxiety*, 28(6):447–455, 2011.

[27] M. R. Morris and D. Morris. Understanding the potential for collaborative search technologies in clinical settings. In *Proceedings of the 3rd international workshop on Collaborative information retrieval*, pages 11–14, 2011.

[28] M. R. Morris, J. Teevan, and K. Panovich. What do people ask their social networks, and why?: a survey study of status message q&a behavior. In *Proceedings of the 28th Annual SIGCHI conference on Human factors in computing systems*, CHI'10, pages 1739–1748. ACM, 2010.

[29] M. Naaman, J. Boase, and C.-H. Lai. Is it really about me?: Message content in social awareness streams. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 189–192, New York, NY, USA, 2010. ACM.

[30] S. Nolen-Hoeksema. *Sex differences in depression.* Stanford University Press, 1990.

[31] I. Pantic, A. Damjanovic, J. Todorovic, D. Topalovic, D. Bojovic-Jovic, S. Ristic, and S. Pantic. Association between online social networking and depression in high school students: behavioral physiology viewpoint. *Psychiatria Danubina*, 24(1):90–93, 2012.

[32] M. Park, D. W. McDonald, and M. Cha. Perception differences between the depressed and non-depressed users in twitter. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, ICWSM'13, pages 476–485, 2013.

[33] S. A. Paul, L. Hong, and E. H. Chi. Is twitter a good place for asking questions? a characterization study. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, ICWSM'11, pages 578–581, 2011.

[34] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1):547–577, 2003.

[35] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599. Association for Computational Linguistics, 2011.

[36] S. Rude, E.-M. Gortner, and J. Pennebaker. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133, 2004.

[37] D. Scanfeld, V. Scanfeld, and E. L. Larson. Dissemination of health information through social networks: Twitter and antibiotics. *American journal of infection control*, 38(3):182–188, 2010.

[38] Statistics-Canada. Individual and household internet use, 2010.

[39] S. J. Sullivan, A. G. Schneiders, C.-W. Cheang, E. Kitto, H. Lee, J. Redhead, S. Ward, O. H. Ahmed, and P. R. McCrory. 'what's happening?' a content analysis of concussion-related traffic on twitter. *British journal of sports medicine*, 46(4):258–263, 2012.

[40] T. Weinberg. *The new community rules: Marketing on the social web.* Oreilly & Associates Incorporated, 2009.

[41] R. W. White and E. Horvitz. Cyberchondria: studies of the escalation of medical concerns in web search. *ACM Transactions on Information Systems (TOIS)*, 27(4):23, 2009.

[42] WHO. World suicide prevention day, 2012.

[43] Y. Zhang. Consumer health information searching process in real life settings. *Proceedings of the American Society for Information Science and Technology*, 49(1):1–10, 2012.

Table 3: Appendix A: Full Comparison of Second-Order Themes, using default LIWC categories.

LIWC Category	Example Terms	Info Dissem. (%)	Self Disclosure (%)	Social Engage. (%)	Range (%)	Mean (%)	Rank
Function words							
Total pronouns	I, them, itself	5.75	16.71	13.66	10.96	12.04	2
Personal pronouns	I, them, her	3.15	12.95	9.01	9.80	8.37	5
1st pers singular	I, me, mine	0.83	11.89	4.02	11.06	5.58	7
1st pers plural	We, us, our	0.10	0.14	0.37	0.27	0.20	56
2nd person	You, your, thou	2.07	0.66	3.38	2.72	2.04	22
3rd pers singular	She, her, him	0.10	0.16	0.73	0.64	0.33	50
3rd pers plural	They, their, they'd	0.05	0.10	0.50	0.46	0.22	54
Impersonal pronouns	It, it's, those	2.60	3.76	4.66	2.05	3.67	19
Articles	A, an, the	2.96	3.89	3.45	0.92	3.43	23
Common verbs	Walk, went, see	7.39	16.90	16.36	9.51	13.55	1
Auxiliary verbs	Am, will, have	3.84	10.29	9.20	6.45	7.78	9
Past tense	Went, ran, had	0.42	2.10	1.71	1.68	1.41	32
Present tense	Is, does, hear	4.83	11.80	12.28	7.45	9.64	6
Future tense	Will, gonna	0.39	1.12	0.78	0.73	0.76	45
Adverbs	Very, really, quickly	2.16	5.03	4.48	2.86	3.89	18
Prepositions	To, with, above	9.77	11.46	8.72	2.74	9.99	12
Conjunctions	And, but, whereas	3.70	4.47	4.42	0.77	4.20	20
Negations	No, not, never	0.79	2.25	3.04	2.25	2.03	25
Quantifiers	Few, many, much	2.32	1.82	1.87	0.50	2.00	35
Numbers	Second, thousand	0.16	0.27	0.23	0.10	0.22	61
Swear words	Damn, piss, fuck	0.03	0.48	0.98	0.95	0.50	46
Psychological Processes							
Social processes	Mate, talk, they, child	7.51	3.88	9.73	5.85	7.04	11
Family	Daughter, husband, aunt	0.17	0.37	0.29	0.20	0.28	57
Friends	Buddy, friend, neighbor	0.07	0.03	0.21	0.18	0.11	62
Humans	Adult, baby, boy	1.08	0.49	1.24	0.75	0.94	43
Affective processes	Happy, cried, abandon	16.90	15.01	18.04	3.03	16.65	4
Positive emotion	Love, nice, sweet	5.16	2.46	4.61	2.70	4.08	17
Negative emotion	Hurt, ugly, nasty	11.40	12.45	13.21	1.82	12.35	10
Anxiety	Worried, fearful, nervous	2.71	0.70	0.69	2.02	1.37	31
Anger	Hate, kill, annoyed	0.64	1.21	2.10	1.46	1.32	33
Sadness	Crying, grief, sad	7.60	9.28	9.27	1.69	8.72	13
Cognitive processes	cause, know, ought	12.72	13.00	14.77	2.05	13.50	8
Insight	think, know, consider	1.93	1.86	2.21	0.35	2.00	36
Causation	because, effect, hence	2.72	1.19	1.83	1.53	1.92	30
Discrepancy	should, would, could	1.17	1.44	1.39	0.27	1.33	44
Tentative	maybe, perhaps, guess	1.38	1.98	1.78	0.60	1.71	38
Certainty	always, never	0.75	1.17	1.84	1.09	1.25	37
Inhibition	block, constrain, stop	0.79	0.32	0.64	0.47	0.58	49
Inclusive	And, with, include	3.25	3.89	3.21	0.68	3.45	27
Exclusive	But, without, exclude	1.51	2.48	2.76	1.25	2.25	29
Perceptual processes	Observing, heard, feeling	0.89	2.92	2.92	2.03	2.25	24
See	View, saw, seen	0.26	0.51	0.63	0.37	0.47	52
Hear	Listen, hearing	0.29	1.44	1.18	1.15	0.97	40
Feel	Feels, touch	0.21	0.93	0.92	0.72	0.69	47
Biological processes	Eat, blood, pain	6.44	2.24	3.73	4.19	4.14	15
Body	Cheek, hands, spit	1.52	0.39	0.52	1.13	0.81	41
Health	Clinic, flu, pill	3.78	1.19	1.54	2.59	2.17	21
Sexual	Horny, love, incest	0.95	0.39	1.29	0.90	0.87	42
Ingestion	Dish, eat, pizza	0.57	0.42	0.57	0.15	0.52	53
Relativity	Area, bend, exit, stop	7.88	16.93	10.61	9.05	11.81	3
Motion	Arrive, car, go	1.38	3.16	1.81	1.78	2.12	28
Space	Down, in, thin	3.60	6.25	3.43	2.82	4.43	16
Time	End, until, season	2.75	6.67	5.26	3.92	4.90	14
Personal Concerns							
Work	Job, majors, xerox	3.00	0.43	1.32	2.57	1.58	26
Achievement	Earn, hero, win	2.20	0.95	1.06	1.25	1.40	34
Leisure	Cook, chat, movie	1.66	1.98	1.85	0.33	1.83	39
Home	Apartment, kitchen, family	0.18	0.32	0.21	0.14	0.24	59
Money	Audit, cash, owe	0.30	0.17	0.24	0.12	0.23	60
Religion	Altar, church, mosque	0.17	0.14	0.37	0.23	0.23	58
Death	Bury, coffin, kill	0.31	0.37	0.51	0.20	0.40	55
Spoken categories							
Assent	Agree, OK, yes	0.30	0.67	0.96	0.66	0.65	48
Nonfluencies	Er, hm, umm	0.10	0.11	0.10	0.01	0.10	63
Fillers	Blah, Imean, youknow	0.05	0.33	0.61	0.56	0.33	51