Online Support

Investigating the role of public online forums in mental health

Josh Smith
Jamie Bartlett
David Buck
Matthew Honeyman

DEMOS

Open Access. Some rights reserved.

PARTNERS CREDITS

CONTENTS

INTRODUCTION

This report seeks to investigate the large amounts of mental health related discussion taking place on public online forums, and to explore the potential for the use of computational techniques to provide robust, actionable insight from these conversations to a wide audience - from healthcare professionals and policymakers to those affected by mental health issues themselves. Furthermore, this study addresses the technical and ethical challenges posed by the collection and analysis of online forum data.

The report describes the collection and analysis of one million publicly available forum posts, collected from forums related to mental health. It details investigations into a number of specific areas: an examination of posts from users asking for urgent help online, public discussion around Cognitive Behavioural Therapy, and comments by mental health patients suffering from specific physical conditions. The report also includes a number of posts, paraphrased for the sake of user privacy, which illustrate the types of discourse which takes place on these forums. The methods used to conduct this analysis are described in a detailed methodological annex.

The last 10 years have witnessed a dramatic increase in the number of people going online to seek out information and advice, to research health issues and connect with those sharing their experiences. NHS Choices, for example, has seen a steady increase in usage every year since 2011, with over 31 million unique visitors to the site in January 2015 alone.

In many ways, the increasing numbers of people accessing health information online should come as no surprise. The Internet offers a fast, free method to seek medical advice without having to make and keep an appointment with a GP. It also provides a degree of anonymity. The 'disinhibition' effect of posting online, often under a pseudonym, has been long documented, and people sometimes feel able to disclose facts on web forums which they might feel uncomfortable discussing face-to-face. Online communities also offer people a way to find and access peer-to-peer support.

Both opportunities and challenges are presented by this shifting landscape. There is a growing gap between changing public behaviour and expectations and the way most NHS services are currently delivered.[1] The House of Commons Health Committee's 2014 report, for

example, highlighted the need for mental health services to keep up with the challenges brought by the Internet and social media, which include the proliferation of online bullying and the advent of online forums or communities that promote resistance to mainstream health advice - for example, 'pro-anorexia' sites. Other research has noted growing fears over inaccurate health information being shared and acted upon, and the dangers of using software to identify and contact people on the basis of what they post.[2]

Equally, there are many examples of online communities being created which constitute valuable online spaces in which people can share experiences. These include well-established Twitter chats. By including certain 'hashtags' in their messages, Twitter users are able to ask questions and share advice with an interested audience. Some of these discussions involve health professionals as key participants - for example, the 'Mental Health Nurse Chat', which uses the hashtag #MHNurseChat, encompasses a wide-ranging discussion, and is employed by the public to ask nurses for advice on mental health issues, and for health professionals to share recent medical papers and news of conferences. Other hashtags cater for specific conditions, and include '#BPDchat', a popular discussion for those with Borderline Personality Disorder.

Some online projects have explored the potential for blogging as a powerful tool for those with long-term health conditions to get their voices heard, as well as acting as a sort of therapy itself. One such project, 'A Day in the Life[3]', provided a platform for those suffering from mental health problems to share their everyday experiences, which were posted unedited and publicly, giving users a valuable opportunity to commit their lived experiences of healthcare, society and public services to the record. Other studies of online support groups and forums shows that people often interact with them for a short time, and that they are likely to lead to improved knowledge and reductions in anxiety, helping people feel less alone.[4]

This report focuses on one particular aspect of this new environment: the use of online web forums. These are similar to 'bulletin boards' - spaces on which people can gather to discuss their interests, seek advice from those with shared experience, and often simply exchange the chatty pleasantries which constitute everyday life. Typically, these communities will spring up around a specific topic, and a wide range of forums exist which allow users to discuss mental health issues, usually in anonymity. While some forums are private, visible only to those who create an account, many are freely and publicly available.

Activity within these online communities has generated an extremely large body of publicly accessible data on health, from discussions on the unexpected side-effects of drugs, to detailed patient opinions on hospitals and forms of treatment. This information is created and shared on a broad variety of sites, ranging from official, accredited sources of medical advice to small, community run forums. These forums potentially provide researchers with a massive, largely untapped qualitative data-source which is incredibly rich, but difficult to access and use.

Online forum data has a number of features which make it unlike other sources of data on healthcare, and health researchers are starting to examine the possibilities which analysing it presents. Data from forums illustrates a "third space", allowing us to understand the lived experience of illness, peer advice and support, and interactions with services. It is created in a novel way, with the perceived anonymity of the Internet facilitating the sharing of normally withheld, or even pre-conscious facets of illness. It allows people to share experiences – which may not be shared elsewhere – of care from organisations and consultations with medical professionals. Forums also allow participants to raise health concerns in 'real time,' as they experience them in the context of their own lives; a record of how participants "do" their illness, rather than how they "see" or reconstruct it in the [often contrived setting of a research interview](), or indeed in the professionalised setting of a hospital ward or GP's surgery. What's more, compared to conventional interviews, forum data is also relatively accessible, and inexpensive to collect at scale.

Scale, of course, presents an analytical challenge of its own. Data from web forums is noisy, imperfect, and generated at a rate which would overwhelm traditional methods of analysis. Computational techniques have been developed, however, which allow researchers to gain insight into these large datasets in their entirety. Machine Learning, a type of Artificial Intelligence, allows researchers to iteratively 'teach' algorithms how to make decisions about a set of data, without these decisions having been explicitly programmed. There is also the field of Natural Language Processing (or 'NLP') which involves the development of methods to allow computers to better process, generate and manipulate human languages. By using these techniques together, it is possible to train automatic systems to make useful decisions about complex linguistic documents.

A number of recent studies have experimented with the use of these techniques to access and analyse similar data sets. In 2015, a *Lancet* study used machine learning technology to analyse medical records of people with psychosis and examine their cannabis use.[5] One 2010 paper

used NLP to sift through posts in health related social networks, looking for reports of adverse drug reactions. It found correlations between frequency of reaction on these forums and the frequency of documented adverse drug reactions. The paper concluded that there are significant challenges, and the field merits further exploration.[6] Further studies have found some benefit in using NLP to extract valuable data from the unstructured (i.e. 'free text') sections of nursing reports.[7]

Online forum data, then, represents a potentially valuable store of information, which remains broadly untapped by the healthcare profession. But the process of gaining any useful insight from these data for health professionals, commissioners, researchers, and patients is difficult. While sophisticated tools have been developed to enable meaningful analysis of large data sets, these have been most eagerly taken up by those with commercial or political interests. In addition to technical and methodological challenges, there are also major ethical and professional considerations of using this type of data, which need to be carefully addressed. These include protecting highly vulnerable participants from harm, and acknowledging that while forums may be in the 'public' domain, the expectation of privacy amongst users may be high. With a small number of exceptions, such as those noted above, little research has yet been conducted into the potential use of these technologies to make sense of the mass of unstructured public health data.

This research paper focuses on the potential of user generated data in unregulated (i.e. not run by health professionals) public and open health forums to provide useable, robust insight for health professionals. It is presented as an exploratory pilot study looking at the potential for this work in the area of mental health, rather than a comprehensive analysis of the subject. It should be seen as sketching out the ways forward for future research rather than providing definitive answers.

To conduct this study, researchers collected 1,070,469 forum posts from six forums related to mental health, sent during a twelve-year period from June 2004 to May 2016. All of these posts were visible to the public, and the forums selected did not require a username or password to access. Different types of analysis were then applied to this data, with the aim of determining what sort of information was available, and which techniques could usefully be applied.

Our exploratory analysis asked three key questions of this data. First, to what extent can we accurately identify "cries for help" - cases in which a user has posted an urgent appeal for assistance? Furthermore, of users

who are seeking help, can we determine whether they have previously sought assistance through the UK medical system? The ability to make this second distinction may help understand trends in mental health, and provide insight into relationships with services.

Second, given the policy priority of Cognitive Behavioural Therapy (CBT), can we identify discussions involving CBT and, within these, work out how it is being discussed and used by patients? This could provide a way of measuring the effectiveness of this treatment, as well as gauging the demand and perception of CBT amongst those grappling with mental health issues. Finally, since mental and physical health are strongly correlated, can we identify and understand how forum posters experience a range of co-morbid conditions (respiratory problems, diabetes and musculo-skeletal problems) alongside mental health issues?

The paper concludes with some thoughts about possible future uses and developments of this approach in mental and physical health, technical issues including the further development of software and tools and the ethical issues of approaches using open, but often highly personal, unstructured health data for analysis.

Given the exploratory nature of this research and its focus on the transferability (or otherwise) of existing machine learning techniques to unstructured health data, this paper sets out in detail our technical methodology and the processes we followed to conduct the research. For the purposes of readability, we have included a short methodology section to accompany the results, with a full methodology in the Annex. We recommend that anyone interested in the approach taken and technical considerations of undertaking this type of work read both.

METHODOLOGY AND RESULTS

This project pursued a multi-phase research methodology, detailed in full in Annex 1.

First, we identified a series of forums from which we could collect data. After consultation with a steering group assembled from health professionals, policy experts and those working in online health, we selected six UK focussed mental health forums. These represented a range of sizes and subjects, from large generalist forums with a mental health area to small communities dedicated to a single issue or condition. Throughout this paper, we have removed the names and URLs of the forums themselves, and will refer to them with a short description. A fuller description of these forums is given in the annex.

We then used web-scraping technology - a technique whereby web pages are 'read' by a piece of software and saved into a database - to collect a large number of publicly available posts from these forums. The privacy of users was a key concern here. While it is difficult to automatically anonymise data, since identifying information is occasionally present in the text of the posts themselves, no demographic or locational geographical data was collected or inferred during this project, and usernames were removed at source.

This collection resulted in the following data set, containing 1,070,469 posts in total.

Table 1

Site	Date of collection	Unique posts	Unique users
Patient advice forum	05/2015 – 06/2016	362,623	24,746
Mental health forum 1	09/2005 – 06/2016	339,752	9475
Mental health forum 2	10/2007 – 06/2016	192,101	4239
Carer's forum	02/2007 – 06/2016	106,656	2495
Mental health medication forum	12/2010 – 06/2016	56,040	2744
Depression forum	06/2004 – 06/2016	13,297	3781

To analyse the data, we built a bespoke dashboard using the analytics package Qlik. This gave us an initial view into the data, and allowed us to determine key themes and patterns of use on these forums. This dashboard was then presented at a half-day workshop at The King's Fund, involving likely end users of this type of software, including

representatives from the NHS and various data science organisations. Participants discussed the usefulness of this data, the potential of NLP to gain insight from it, and likely ethical considerations.

Researchers then used a piece of software named 'Method52', developed by Demos in partnership with the University of Sussex, to train a number of NLP classifiers to search for patterns within the data. In doing so we focused on answering three specific questions, arising both from our use of the dashboard and the responses from the workshop.

Firstly, since one of the goals of this research was to understand the types of discussion underway on online health forums, we wanted to understand whether we could identify people posting 'cries for help' - using the forum to reach out to a community for urgent health advice - and whether NLP classifiers could be trained to automatically identify these posts. Secondly, since Cognitive Behavioural Therapy (CBT) was identified during the workshop as a current policy priority for the Department of Health, we investigated whether NLP could be used to identify and characterise the conversation around this treatment. Finally, we investigated the prevalence of conversations around conditions with a high incidence of co-morbidity.

The results of this research, as well as a general description of the dataset, are set out below.

Data, volumes and users

As part of the web-scraping process, researchers used user information presented with forum posts to assign anonymised user IDs to the data. Due to the differing ways in which forums are constructed, it was not always possible to assign a user ID to each post. Accordingly, IDs were assigned to 451,000 of 1 million posts collected.

The forums chosen for this project differed radically in terms of size and patterns of use. The largest site in terms of activity and user base was the patient advice forum, containing around 360,000 posts from 25,000 unique users, while the smallest, focussing on medication, contained only around 13,000 posts.

In order to better understand the nature of user activity in these forums, we calculated the average number of forum posts per user, which highlights some stark differences in the amount of activity. In particular, some of the smaller forums showed a high number of posts per user, possibly suggesting that a community or regularly engaged users has

built up on these platforms. This could also be due to differences in forum focus – the forum related to patient advice, which is primarily a message board for those seeking medical advice on specific conditions, saw far fewer posts per user than the carer's forum, a forum for those caring for people with long term conditions, which is likely to serve more of a social purpose.

Table 2

Forum	Average posts per user
Mental health forum 2	45.3
Carer's forum	42.66
Mental health forum 1	35.87
Mental health medication forum	20.45
Patient advice forum	14.65
Depression forum	3.51

The dataset contained a high number of posts sent by a small number of prolific users. Of those posts in the dataset which had a User ID attached, half of were sent by only 2 per cent (501) of identified users. Initially, researchers posited that many of these prolific users were likely to be generic accounts (e.g. 'Guest') signifying users who have not fully registered for their accounts on a given site. Due to our anonymisation of users, it was difficult to verify this at scale. However, researchers manually investigated the first 15 most active accounts, and found that all but one of these was an account appearing to belong to a single individual.

Links and mentions of other sites and organisations

In order to characterise the topics discussed on mental health forums, researchers used a technique called 'Named Entity Recognition', described in detail in Annex 1. This allowed us to generate a list of terms present in the dataset which were likely to refer to organisations, presented in Table 3 below.

Table 3

Organisation	# mentions (by NER)
NHS	5332
Social Services	1626
Department of Work and Pensions	1164
BBC	1018
Food and Drug Administration	951

In order to generate the list above, it was necessary to manually 'clean' the terms identified, and we have removed non-organisational words such as 'GMT' and 'Hi", leaving all and only terms judged by researchers to properly refer to an organisation.

The list above suggests that a significant amount of discussion on these forums involves formal services – the NHS, for example. Accordingly, health forums could be a useful resource for further analysis of service delivery, although it is to be noted that people do not always know from which organisation they are receiving support.

Of particular note is the prevalence here of the Food and Drug Administration, which is a US based organisation responsible for the regulation of drugs. This illustrates that these forums, although based in the UK, also play host to conversations about other countries. While we explore techniques to minimise this below, it may not be possible to entirely remove this 'global' conversation from a dataset, and any study involving analysis of data solely concerning the UK will need to take this into account.

We also identified the most referenced links in the data set. The top four of these were all popular social media platforms (Facebook, YouTube, Twitter, and Google Plus), which suggests that these forums play an important role in a much broader 'social' community of users and content. This data also allows us to identify popular resources shared by those dealing with mental health issues - for example, 'Little Porcupine Goes to the Psych Ward,' a novel examining one woman's experience with the American medical system.

Table 4

Link to third party website	Number
www.facebook.com	16902
www.youtube.com	14799
twitter.com	11389
plus.google.com	11104
digg.com	10971
www.reddit.com	10886
www.stumbleupon.com	10765
patient.uservoice.com	9293
play.google.com	4916
cepuk.org	4064
itunes.apple.com	3716
paper.li	3490
en.wikipedia.org	3194
beyondmeds.com	2311
www.littleporcupinegoestothepsycheward.com	2040

Natural Language Processing classifiers and case studies

The first two lines of investigation followed in this project - identifying cries for help and mentions of Cognitive Behavioural Therapy - involved the building of Natural Language Processing (NLP) classifiers. The use of this technique responds to a general challenge of social media research: the volume of data routinely produced and collected is too large to be manually classified. As set out in the methodology section, NLP classifiers provide an analytical window into these kinds of datasets. They are trained by analysts on a sample of the data, and are designed to use linguistic patterns in documents - in this case, forum posts - to distinguish documents based on their meaning. This training is conducted using a

platform called 'Method52', developed by Demos in partnership with the University of Sussex to allow non-technical analysts to train and use classifiers. This process is discussed in detail in annex 1.

Investigation 1: Cries for help

One potentially positive aspect of online health forums is that they are highly available; users are able to post from their own home, at any time. This enables them to act as a means of support in times of crisis, when traditional help – through a GP, for example, or a therapist – isn't available. We wanted to explore the potential for NLP to aid with identifying posts that were high-priority, urgent requests for help or advice. To do so, we built a three-stage pipeline, detailed below.

Firstly, in order to increase the likelihood of finding people requesting help from other forum members, we first identified all posts which were the very first posts in a thread, and removed others from the dataset.

Secondly, we built a classifier that aimed to sort these first posts into two distinct categories: 'needs help' and 'does not need help'. 'Needs help' was defined as those containing requests for help, including; users seeking mental health advice, for themselves or those they care for; users seeking advice on how to access care; users seeking advice on types of care; users who are 'at the end of their tethers' and turning to the forums for support and; users who mention needing an urgent change, but who don't know how to make it. For example, a post which included the phrase 'I am hoping to find support here and new friends' would be classed as 'needs help'.

Posts judged not to be urgent requests for help included users who were undergoing treatment and were broadly happy with it, users soliciting non-urgent advice, users sharing their stories and experiences in a conversational way, and all other posts.

This classifier identified 95,544 posts as 'cries for help', representing 60% of 158,548 identified first posts. The classifier achieved an overall accuracy of around 65 per cent. (Classifier results and a discussion around difficulties encountered while building them are discussed in Annex 1.)

While this classifier was not able to reach overly high levels of accuracy, it was judged accurate enough to enable us to create a filtered dataset of first posts asking for help, which we then classified further.

Users identified as asking for help often mentioned contact with health professionals. If these posts could be identified, this data could potentially be used to flag up failing services and common complaints, or to identify areas of care which people in need felt they were missing. Accordingly, a classifier was built which attempted to identify, for each post classified as asking for help, whether the user posting had received professional medical treatment or therapy. This classifier sorted posts into three distinct categories, as follows. 'Sought help': posts which explicitly mention seeking help from a medical professional, 'Not sought help': posts which do not mention seeking professional help, and 'other': posts which do not mention a personal problem. An extract from an example of a post identified as 'not sought' is below (the wording of all posts here has been altered to preserve user's privacy).

> *"...I had to finish in my job due to being off 6 months. I'm worried about my rent due to my benefits not being in, I'm worried I won't be able to afford christmas. I want to self harm as it helps release the thoughts and pain and anger in my head, I'm trying to keep my mind out of the state but I feel that I'm at the end now and want to die. I guess I'm just feeling lonely, worried, and don't know what to do."[8]*

While building this classifier, researchers came across a number of notable edge cases – for example, users who mentioned treatment for a problem unrelated to the one causing the current crisis. One such user talks about visiting a GP in order to seek help for depression, but feeling unable to ask:

> *"...i just dont know how to cope for the whole week i havent slept more than 8 hours total for the whole week so im exhausted. i went to my gp last week to ask for help but i i couldnt tell her, i chickened out, just got a repeat prescription instead..."*

This classifier identified 50,533 posts from users who had sought medical help, and 24,132 from users who had not. The classifier achieved a fairly low overall accuracy of 56%, although the label 'sought' was applied with a higher accuracy of 69%.

This investigation shows that there is a significant amount of forum activity from users suffering from mental health issues reaching out to other forum members for urgent assistance, though the precise level of this activity is likely to be slightly overstated above. Due potentially to the wide variety of language used and the prevalence of edge cases, it was difficult to train NLP classifiers making these distinctions to a high level of accuracy. However, researchers did find that passing posts through these classifiers

enabled the creation of a sample containing high numbers of relevant posts, and it is highly likely that this approach could be a valuable method of identifying data for qualitative or thematic analysis.

Investigation 2: Cognitive behavioural therapy

One potential use of our approach to forum data is to identify patients' experiences with specific drugs and therapies. During phase I of the research project and the workshop held at The King's Fund, it was suggested that we try to identify posts mentioning experiences with Cognitive Behavioural Therapy. CBT is a type of talking therapy that can be used to help people with a host of psychological problems, but it is particularly recommended for the treatment of depression and anxiety. It has been increasingly prescribed on the NHS over the last ten years as part of the national Improving Access to Psychological Therapies (IAPT) programme, or as part of services offered by GPs and community mental health services.

We were interested to see whether we could gauge demand for the therapy from the posts we'd collected, and find users who were interested in undergoing it but had not. We were also interested in identifying patients' experiences accessing the therapy, and their accounts of the perceived benefits or drawbacks of the treatment.

The first stage here was to filter out posts which did not mention CBT. To achieve this, researchers used Method52 to apply a 'keyword annotator' to each post in the dataset. This labelled all posts containing a given list of terms, allowing us to remove posts which did not contain at least one of the terms 'CBT', 'Cognitive Behavioural Therapy' or 'Exposure Therapy', a variety of CBT.

We then built a classifier only using posts containing one of these terms. For each post, it aimed to distinguish whether the post mentioned a personal experience of the therapy. The classifier split posts into two categories: 'have had', which included all mentions of people who appeared to be discussing a personal experience, or who mentioned current or future treatment, and 'other', which included all posts which did not mention a personal experience with CBT. This second category included users recommending the therapy without mentioning their experience, discussing its effectiveness, or asking what it is.

Of 8374 posts which mentioned CBT, 5582 (66%) were classified as 'have had'. This classifier was more accurate than those built during the first investigation, achieving an overall accuracy score of 65%, with a high

accuracy (72%) for the 'have-had' label, and an accuracy of 53% for 'other'.

Many posts classified under the 'have had' label contain a deal of interesting information on self-care, including descriptions in personal terms of the way in which CBT has affected patients, and discussion of medications, apps or websites which helped:

> *"...I thought thinking about my issues (essentially why I was like this, what was wrong with me etc.) would eventually offer me an answer to overcoming my depression. I guess that's why it was a never ending cycle. Practising mindfulness and CBT has helped me to improve though..."*

> *"...Thanks - I'll keep the antihistamine in mind. I have tried Nytol, which is antihistamine based and made me feel awful, but didn't help me sleep. Maybe it was too strong. I've been considering taking the Sleepio course which is an online CBT course with lots of good reviews, so might give that a go if things don't settle soon..."*

Our findings suggest that this approach, combined with a keyword search for the therapy in question, may prove fruitful to researchers aiming to identify public conversations about a particular treatment or medication.

Investigation 3: Co-morbidity

Many people live with long-term physical conditions alongside mental health problems. When they do, they are more likely to suffer worse health outcomes for their physical health than those living without a mental health problem.[9] We wanted to identify user discussion around such co-morbidities, as well as determining whether a simple keyword search, rather than an NLP classifier, would yield relevant results.

In each case, we built a list of relevant keywords, and randomly selected 50 posts which contained one of them. These samples were then manually sorted into categories.

Study 1: Respiratory disease

Researchers used a keyword annotator to identify all posts containing the terms 'asthma' and 'COPD', yielding 2655 posts from the total collection of 1 million. In order to identify users mentioning mental health problems and respiratory conditions (indicating a co-morbidity situation),

we then filtered out results from the 'patient advice' and 'carer's' forums, which were not explicitly focussed on mental health. This left 366 posts of the original set of 2655, from which a random sample of 50 posts was chosen.

This method yielded a high proportion of posts relevant to asthma and mental health co-morbidities. Of the 50 posts we reviewed, only one was judged to be irrelevant suggesting an accuracy score of approximately 98%, the highest accuracy score amongst the co-morbidities.

Within this data set there were a number of posts which could be of value to researchers. One post, for example, contained a very vivid example of a patient's experience of the mental health system, which disregarded his physical long-term condition (asthma) and vice versa, though it's not immediately clear which mental health conditions they suffer from:

> "I don't know how to get this message to the relevant people; could you please forward it for me? I've met a large amount of people in hospital who have suffered similar problems. I have been in hospital and they've taken away all my medication, but then failed to give me my normal prescriptions, in particular, for physical illness. I have often gone without asthma and also allergy medication for days in a row. When I have been on a Ward for physical illness I've gone without my usual mental health meds, each time saying that the pharmacy has not yet sent them down to the ward. I've told them going without my daily meds could have dangerous side-effects but they don't seem to understand or care. Also you are expected to take any new pills they prescribe, despite worrying lists of side-effects. I've known people to die suddenly when just out of hospital or even during their stay, with this put down to natural causes, unexpectedly young. Being mentally unwell means I find it difficult to ask for help, hard to complain, hard to voice concerns or opinions. I feel I don't have the right to even be concerned about these things, as I see that without medical intervention I would surely have died by suicide, so this is extra time and I need to accept whatever I get."

Some posts included detailed experiences with particular types of treatment, and how patients managing respiratory conditions alongside a mental health problem:

> "I saw a counsellor some time ago who was into Neuro Linguistic Programing or NLP. In number of sessions he was able to put me in a semi conscious state where I found myself having my first panic attack, driving. He even made me a double sided relaxation tape

which I still use; it's sort of personalised. The most helpful book I have ever read in respects to my kind of complaint (Anxiety, Depression) is Self Help For Your Nerves, it really is useful. It came everywhere with me in the second half of the 80's."

Study 2: Diabetes

This study aimed to pick up discussion around diabetes. Using the same procedures as above, we identified posts containing one or more of the terms 'diabetes', 'insulin', glucose', 'hypoglycaemic' and 'hyperglycaemic'. (We did not distinguish between type 1 and type 2 diabetes). The latter three terms were added as we suspected they would be discussed mainly in the context of diabetes experiences and advice on these health forums.

This yielded a total of 3561 posts. Once again, we filtered posts on forums that weren't specific to mental health, leaving us with 993 posts.

A random sample of 50 of these was found to contain a high proportion of relevant data, with only six were judged to be irrelevant, an accuracy of 88%. All 6 of these irrelevant posts were academic content relating to glucose, but not specifically related to diabetes.

Most of the posts contained discussion of diabetic issues, including patients seeking advice regarding their conditions, strategies for managing diabetes (including weight loss), or posting about symptoms they were experiencing which made them wonder if they were diabetic. We also found users sympathising with others who were diabetic, or discussing their own experience of living with the condition.

A number of posts noted the relationship between anti-depressant medication and diabetes:

> *"Every psychiatric drug effects blood sugar and can lead to diabetes if you're already predisposed. The worst are neuroleptics, but ADs definitely affect things too. I've previously posted links to lots of additional info that help mitigate the harm for people like us - good luck."*

Study 3: Arthritis and osteoporosis

For the study of MSK, we took a set of candidate keywords, sourced from a WHO paper on the burden of MSK conditions: 'Rheumatoid arthritis', 'Osteoarthritis', 'Arthritis' and 'Osteoporosis'. The list initially included the phrase 'lower back pain', which resulted in a very wide selection. We therefore removed that term and were left with the more technical terms listed above.

The total number of posts from our dataset referring to these MSK conditions produced by our keyword annotator was 8237. Removing posts made on non-mental health specific forums yielded a set of exactly 500 posts. We sampled 50 of the posts which were manually classified, but found that only 7 were relevant to experiences or advice about living with a mental health condition at the same time as a physical health problem, an accuracy of 14%.

Inspecting the posts classified as irrelevant to co-morbidity, the reason for this lower accuracy appeared to be the frequency with which the keyword annotator was picking up posts discussing the many MSK conditions that are possible side effects of medication for mental health problems, rather than users affected by the conditions.

Of the posts that were clearly about living with both types of condition at the same time, some insight can be gathered into the relationship between MSK symptoms and the symptoms of mental health problems:

> *"I believe that physical activity is the greatest antidepressant. Anyone can get a referral to their local sports center or gym if they are associated with the programme and get reduced cost sessions. Regrettably I suffer from severe arthritis, and when it comes back I can't exercise, normally when this happens my depression returns. In my opinion medication alone will never cure depression , you need a positive mentality. Many people find that they don't need medication while exercising , I find it can cause nausea to have lots of medication in your system while training. Definitely speak to your Doctor or Psychiatrist about reducing AD medication while being involved in a physical activity"*

Posts also yielded insight into the experiences of patients with both physical and mental health problems interacting with the health system at large:

> *"I've joined this forum because I feel completely isolated. I know I have a reclusive personality and sometimes get depressed but I'm not suicidal and can normally cope. As a child I had selective mutism. I couldn't speak outside my home but I got no help for it, just criticism. I've never believed that I'm worth the same as other people and don't look after myself in many ways. For example, I've had no medical practitioner for a long time. I used to be registered with a local doctor and had medical care for arthritis and depression. I saw a female doctor at the surgery for HRT treatment and when she moved to a different practice I was surprised to hear I had been struck off her list. I didn't know what to do about it and so I didn't do anything. Earlier today, however, I made a written complaint and have asked for assistance in*

finding a doctor again."

CONCLUSIONS

Our conclusions here focus on the extent to which this type of work can be conducted in future, and are broken into three parts. First, possible future directions. Second, insights into building software for this purpose. Third, principles of research in this area.

Future directions

We have demonstrated that there are enormous volumes of data available (over one million posts, from only six forums) and that this data is amenable to capture and analysis. However, there are some significant technical, methodological and ethical challenges still to overcome, explored below.

Based on this work we think that there is a significant amount of valuable data about current services, conditions, and products present on online health forums. This data is amenable to collection, if not necessarily automated analysis. Some of this data is highly detailed, and is likely to prove valuable as a means to supplement existing feedback data. Furthermore, we feel that named entity recognition and natural language processing are both viable approaches to help identify and classify information that is likely to be of value to professionals, though we have shown it does not function well in all cases.

This is one of the first times that unstructured health data in a complex and nuanced health area has been collected and classified in this way. In the longer term, this proof of concept could help:

- Owners of forums to better understand the topics and issues discussed - how these differ between forums, and why. In the longer term, this may lead to the possibility to offer well-targeted specific information or relationships with service providers to communities of users, as well as membership of a forum being seen as a self-management offer itself).

- Service providers, both in and outside the NHS, to develop a better, deeper and more truthful understanding of users' experience with services, which allows a more thoughtful design in response. NHS service representatives in our workshop saw big potential in this sort of usage, though were also well aware of the pitfalls. In particular, there are likely to be demographic groups who are not being included in online analysis, which could create new forms of inequality (see below).

- Health regulators to access additional insight into the organisation specific or overall performance of different providers.

It is, however, important to note that this approach is not and never will be a silver bullet. It should be seen as an alternative and complementary source of information and meaning, rather than a replacement for other methods, such as administrative health datasets, specific surveys of focus groups, or more specific performance information.

The information in these posts has not been designed to answer specific questions. This lends it a number of strengths, especially in that is a source of unbiased and unguarded full and complex accounts. This is rare in qualitative studies in health, since responses and behaviours are known to be influenced by the knowledge of being observed; a phenomenon known as the Hawthorne effect.[10] But the data's lack of built in focus also means that it has to be interpreted, and this process is context-specific and value-laden. This data, and the meaning flowing from it, is not representative and should not be claimed as such. We do not know whether or not users of health forums are currently representative of the wider population, and more work is necessary to determine further what demographic biases are involved in this type of data, without compromising user privacy.

This project has demonstrated some of the possible benefits of using unstructured health data in the complex and multi-faceted area of mental health. Given this, there are some obvious further areas of study:

- Looking at patterns over time in terms of forum topics - e.g. to pick up leading indicators of changes in incidence of particular health issues, or identify leading indicators of performance issues with providers.

- Understanding the "careers" of forum users. To what extent are these forums a source of long-term ongoing support versus short-term problem resolution? How might patterns in the data be used to develop the use of public forums as part of a self-management offer to targeted groups of patients, by type of issues, or characteristics of posters, or service use? Forums are potentially vital sources of peer-to-peer support that can contribute to user well-being, but further work is required to better understand this process.

- A more detailed examination of the relationship between social media and forum behaviour.

- Discriminating between high-intensity posters and lower-intensity posters to distinguish types of poster and the issues that they use the forum to raise or address, and to what extent high intensity

posters skew overall findings.

- Understanding / gaining insights into the characteristics of successful engagements in order to assist forum owners in designing, developing and sustaining them - we have noted, for example, very significant differences in the activity and engagement across different forums.

- Exploring ways of using forum data to address country or city-specific questions, for example on the quality of care in a specific area, given that topics of discussion on UK specific forums are not limited to the UK.

Building software

When building software for this task there are a number of technological considerations which can cause issues if not accounted for. These include, but are not limited to:

Time

One of the major technological issues faced in this project was one of computation time and power. Performing scrapes of large web-forums can take a considerable amount of time, regardless of the computational resources available. This is due to the necessary delay of anything between 0.3 and 0.5 seconds between calls to a domains server to prevent being blocked as a possible denial of service attack. Secondly, there is often an upper limit of anywhere between 1 and 10 crawlers operating from the same origin (i.e. IP address) that can generally access a site without also being blocked.

Computational power

Having sufficient resources to scrape across multiple domains simultaneously is necessary in order to be able to collect large quantities of data within a tractable amount of time. Within the scope of this study a single quad core desktop machine, with 32Gb of memory, was capable of managing multiple crawlers over up to 5 domains without any technical difficulties.

Robust parsing structures

As web-domains evolve and as the number of auto-generated web-sites increases, the potential for poorly formatted or highly intricate structures within the html markup of web-pages increases. To counter this, it is

advisable to have a number of html parsing technologies which can easily identify markup delimiters, for ease of understanding the structure of content. It is also necessary to build software which can be easily adapted to a number of different online technologies and structures to reduce the amount of additional work needed when attempting to parse content from new web domains.

Permissions

Although it is not always possible, it is advisable to review the relevant terms of service of a website to ascertain what a given sites policy is on automated scraping of content from their domain. In several cases, there will not be any and the minimum precaution taken should be implementing a small amount of delay between calls to the server in question to prevent having the IP address being used from being blocked as a potential denial of service attack. In some cases, a site may choose to block any attempt to scrape their content and this should be respected by ceasing all scraping of that domain. This is true of any social science research, but is especially the case in sensitive research work of this nature. (We refer here simply to the technical considerations; ethical issues are discussed below).

Restricted crawls

Due to the stochastic nature of a web-crawl, a crawl in its most basic form simply follows the in and out-links present in a web page. However, it is possible for the crawl to stray into other domains, and potentially scrape content which is not related to the research being conducted. To counteract this, it is advisable to restrict the crawl to only scrape pages from the domain of interest, effectively penning the crawl into only search for posts within the site it was originally intended. Again, this is extremely important when dealing with potentially sensitive data.

Storage

When crawling large amounts of rich, highly inter-related content it is necessary to ensure that all information and structures are retained as closely as possible within the choice of storage. For example, when scraping forums, it may be necessary to retain the order of a thread.

Principles for future work

In addition to software considerations, research work of this nature is extremely controversial and results in difficult challenges about how to conduct research or insight in this area.

Clarity over data access

So-called 'black boxed' data – provided by a third party without clarity over methods, search terms used, or access levels - should be avoided wherever possible. This means that 'off the shelf' data analytics tools are likely to be less valuable than systems that allow researchers and analysts control over how the system operates. We believe that health services should have dedicated systems and in-house specialists if they decide to conduct work of this kind.

Clarity over sampling methods

Data collected from web forums is more difficult that both traditional polling techniques, and research data collected from, say, Twitter, which is relatively well established. This adds a high degree of uncertainty into the demographic background of any collected data set. In other research work, we have identified the need for more robust systems for selecting sample terms as a key methodological innovation necessary to improve the discipline as a whole.[11] Where automated data collection is used, it is important for researchers to clearly state the level of access to the data set they had; precisely which search terms or seed accounts were used and how those search terms were arrived at and why; and whether these decisions result in any likely systemic bias.

Provide performance indicators of automated systems

Where automated data collection or analysis is used to collect data and to automatically classify them using algorithms into categories (such as 'cries for help' or 'not cries for help'), the performance of classifiers and algorithms should be clearly stated, ideally using precise scores and descriptive text explaining the performance levels. Where possible, algorithms should be benchmarked against what results the researcher may have expected, or how other similar studies have performed.

Longer time scales and interdisciplinary work

The entire discipline of social media research is in its infancy. This means that a longer process of engagement than normal is required to expect

tangible outcomes. The process of re-purposing the software was more time consuming than we had anticipated. Much of the financial support at the moment is best placed in fundamental research development rather than more applied work, and this requires genuine inter-disciplinary collaboration between, for example, computational scientists, social scientists, health professionals and patients / patient groups. One possibly important aspect of this could include discussing the implications of this work with the site administrators or 'run' the site. This also means, in practical terms, timescales for research outputs might be longer than for other types of research work.

Maintaining ethical standards

Conducting research using data from online forums presents ethical challenges in respect of how researchers should collect, store, analyse and present data. Due to the relative novelty of this field of research, there are no widely accepted protocols and approaches for how to do this ethically.

This is further complicated because of the often very sensitive nature of health related data. One member of our advisory group (see annex 1 for more on the role of this group) contacted in the early stages of the project commented that within mental health there is an important emphasis on engaging people in their own recovery and treatment. This should also be extended to the research. Notably, researchers should look for ways to 'give something back', asking themselves 'How might we show participants the results of our research?'. Part of the challenge here is the danger of destroying or undermining confidence in these forums, which are often used by people to look after their own conditions, manage them and get advice.

Another challenge is posed by the 'free text problem', in that unstructured text fields may contain incidental personal data which could be difficult to remove; while it is possible to pseudonymise data, full anonymisation is very difficult.

There is also the question of which ethical framework to use when conducting further research. This project was approved by the University of Sussex Ethics Review Panel, who commented that they did not consider that the current paper should be considered a clinical study, since the research did not recruit patients from the NHS, and did not involve the gathering of clinical data, or the conducting of interventions which would affect the care anyone received. The panel

recommended that this research should be properly reviewed under a traditional research ethics approach.

Generally speaking, the Economic and Social Research Council (ESRC) principles of ethical research are an excellent guide for conducting research of all kinds – and can be usefully applied to online as well as offline research. Social media research should adhere to the research ethics standards set out by the ESRC's principles. The key questions are whether or not the research has sufficiently explained the risks and minimisation strategies for:

1) The potential identification of individuals

2) Whether or not the research has sought informed consent, and, if not, why it is not considered necessary (with reference to the expectation of privacy a research subject might have)

3) Whether there is any possible harm to the individual, and what measures there are to minimise them

4) Whether techniques to 'cloak' or protect the identify of research subjects are necessary, and how that might adversely affect the quality of the research

As a very general principle, where an individual is identifiable, explicit permission should be sought, unless a) it is clear that the subject has no expectation of privacy and b) the research will be significantly adversely affected unless the individual is identified.

More specifically, care must be taken in order to avoid publication of details that might lead to the identification of a user. This might be more than just user names. Many users will use pseudonyms, but in the view of the Sussex ethics panel this is not usually a sufficient form of masking, since pseudonyms are very often easily linked to a user. Care needs to be taken about what to publish, since it might lead to the compromise – and therefore harm – of the forums, which might hinder positive activity taking place there.[12]

ANNEX 1: METHODOLOGY

Phase 1: Identification of case studies and ethics

Online health forums are extremely varied in nature, and carry a number of ethical considerations that needed to be worked out before starting data collection. We created a small advisory group of health and informatics specialists, clinicians and strategic public health leaders who met at the beginning of the project to advise on key areas of focus and risks and opportunities of the project.

To inform our decisions, we also conducted a baseline review of existing research work on unregulated health forums. This included a literature review of existing research on the subject of unregulated health advice online, in order to baseline the research project against current research. This included 10 unstructured interviews with health professionals, patients' groups, clinicians, and experienced users of these sources of information.

After consideration, we decided mental health forums offered the most appropriate case study. They are both voluminous and very varied in the types of forums available. Further, mental health is widely agreed to suffer from poor resourcing in traditional types of care and a significant amount of online activity.[13]

We wanted to examine forums from a range of topics and sizes, from large generalist forums with a mental health area to small communities focussed on a single issue or group. To this end, we worked with The King's Fund to initially identify five forums to collect data from, found through an online search; a sixth was identified later through inspection of links within the initial collection (see below).

In order to conduct this research work ethically, we submitted this project through the University of Sussex Ethics Review Panel. New types of online data collection raise new ethical challenges in respect of how researchers should collect, store, analyse and present publicly posted data.

We concluded that we could only collect data from forums which were open (i.e. where no restrictions whatsoever, either human or automated, had been put in place to prevent data collection); and we used an ID hashing system to turn all user names into unique identification numbers to protect the identity of the user – even for the researchers. A hash value for any given forum member was based on the specific string of

characters which comprise the members' user name. Each hash is calculated as a function of each individual character's numerical representation and position within the string, raised to an arbitrary prime number to ensure that each hash is unique for each member. We then 'masked' any quotes in this report, by altering wording while preserving meaning, and removed identifying information so individual users could not be identified on publication. This masking process was tested for effectiveness by running altered quotes through a plagiarism detection system, which ensured the original posts could not be found through a search engine.

Phase 2: Technology planning

This project employed two types of technology, both of which draw on pre-existing software and research developed by Demos & University of Sussex. First, web-crawling to collect the data. Second, NLP to analyse the data.

Web-crawling is an automated process which is used to find and catalogue information stored on websites. Starting with an initial set of seed webpage addresses, a crawler follows the links found on those pages, discovering new webpages which, in turn, give rise to additional links that can also be followed. It is common to place a bound on the number of links away from an initial seed webpage that should be explored. In this way, a webcrawl discovers a collection of potentially relevant webpages. The text on the webpages can then be extracted, stored in a database, and indexed so that efficient search can be performed. (In this work, we did not crawl the world wide web, but rather used the same technique in order to identify and then scrape data from web pages on the pre-identified forums).

NLP is long-established sub-field of artificial intelligence research. Our study makes use of a web-hosted software platform, developed by the project team, called Method52.[14] Method52 uses NLP technology to allow the researcher to rapidly construct bespoke classifiers to sort defined bodies of text into categories (defined by the analyst). Classifiers are algorithms that automatically place text in one of a number of pre-defined categories of meaning. The process of creating a classifier with Method52 is achieved through 'markup'. Text is presented to the analyst via an interface. The analyst reads each text, and decides which, of a number of pre-assigned categories it should belong to. The machine learning algorithm looks for statistical correlations between the language used and analyst's markup to derive an association between the features of the language and the categories of meaning. Having

learned these associations, the computer applies this criteria to additional (and unseen) text and categorises them along the same, inferred, lines as the examples it has been given. This provides a way for a non-technical analyst to productively engage with large textual data sets, supporting a process that involves discovering where scope for (often unforeseen) insight lies within the dataset, and picking apart the dataset in a way that is able to effectively deliver that insight.   Further details of Method52 can be found in Wibberley et al. (2014).[15]

Both of these technologies have long-standing use in a number of research and commercial settings, and both have been developed by Demos and University of Sussex for use on large scale social media data sets. However, as we set out above, they have not been applied to online health advice to the same extent. Therefore, the research team needed to repurpose the web-crawling software, so that it could be used equally effectively on static websites, forums, blogs and peer support groups.

The functionality available within a standard web-crawler was not entirely sufficient to fulfil the tasks required within the scope of this project. Apache Nutch, an open-source project designed for large scale web-crawling, coupled with the Apache Solr search-engine, provided a highly extensible and configurable framework, enabling researchers to begin work. The additional features of the technology were as follows:[16]

Generalising across different forum technologies

Extracting posts from an online forum can be extremely difficult when aggregating over different sites, due to the heterogeneous design and format of html across different web-domains. When attempting to parse these sites we needed to be able to quickly and reliably build web-scrapers, designed to extract information from a specific web-domain. One feature was to extend the features of Nutch to enable researchers to quickly and programmatically describe the specifics of any given forum technology, using two basic high level concepts of a page and an individual forum post (contained within the page). Following this the subsequent features of the framework are able to perform further parsing of a crawled page and store this information with the relevant fields of the search-engine implementation using Solr, effectively homogenising the retrieved data across a variety of places on the web.

Forum parsing

To begin exploring the documents for information regarding forum posts, the content of each web-page was first parsed into a data structure which could be read by our software and deconstructed programmatically according to its html markup. In addition to parsing page content and forums posts we needed to parse for additional information surrounding the individual posts or pages of a forum. For example, this included retaining the structure of individual threads of discussion, anonymised member information, the date of posts and any web links mentioned within individual posts. In order to accomplish this, the scraped web pages and its posts were passed through an annotation framework to add meta-data which was subsequently stored in the search engine.

Identifying additional forums from an existing collection

While standard search engines can be a valuable tool for researchers wanting to discover health forums and communities online, it is likely that by using this method alone a significant portion of forums would be overlooked. Each search engine applies its own algorithms for choosing which results to display most prominently; for example, sites might be ranked on the number of links from other sites which refer to them, or the regularity with which they are updated. Accordingly, smaller, less well-networked sites are often difficult to find through these methods.

After collecting structured data from the sites identified through search engines, we decided to investigate an alternative method for discovering forums. To do this, we used a 'links' field generated through the web-scraping process which holds, for each forum post, a list of URL links contained within that post's content. These lists were then processed using Python's urllib module, which allowed us to extract the base website which a particular URL referred to – allowing us to assert, for example, that the links *http://www.nhs.uk/Conditions* and *http://www.nhs.uk/Conditions/Pages* both referred to the site www.nhs.uk.

These sites were then ranked in order of the number of times they occurred within the dataset, and links to the sites we were collecting from were removed. A sample of 50 links was then analysed by a researcher, to determine the type of website they referred to.

This method produced a number of relevant sites, including four forums relevant to issues around mental health - none of which were discovered through the initial search engine based research - and six relevant blog sites. Of the four forums, one, involving mental health medication, was

chosen to be included in our collection. Based on this experience, we believe that a link-based approach to discovering forums may prove fruitful in the future.

Structured storage

To ensure that as much of the integral structure of a web-forum was maintained within the search-engine containing the data, we implemented a more sophisticated means of communication between the web-scraping technology of Nutch and the storage capabilities of Solr.

Further data processing with Python

Once the structured data had been collected and indexed within Solr, it was exported to Python dictionary format, enriched using a basic Python script written by researchers at Demos.

This script performed a range of tasks, such as ensuring that character encoding was standardised across forums. It was also used to write a new field into the data, concerning a post's position within a thread. The script then exported the resulting data to CSV format.

Phase 3: Data collection

Before initiating collection, we checked to ensure that there were no relevant restrictions in each forum's robots.txt - a publicly readable configuration file or set of html tags which place certain restrictions on how automated web-crawlers interact with the site.

Phase 4: Dashboard creation and analysis

In order to undertake analysis of the data, we undertook a number of steps.

First, we iteratively built a dashboard using Qlik (a data analytics package) in order to provide analysts with a 'window' into the datasets. In simplest terms, dashboards are single screens that aggregate and display multiple flows of data, allowing researchers to quickly discover patterns and anomalies. Dashboards have become one of the ways that decision makers across many sectors in society now receive and digest information. During this process, a number of prototype dashboards were developed, tested by researchers at Demos and The Kings Fund for usability, and then iteratively improved. Often, this involved making

changes to the data collected and displayed: for example, changing character encoding so that data could be read properly, or undertaking deeper or more specific scrapes of data from certain sites.

Fig. 1: final dashboard used in workshop (Identifying details have been obscured)
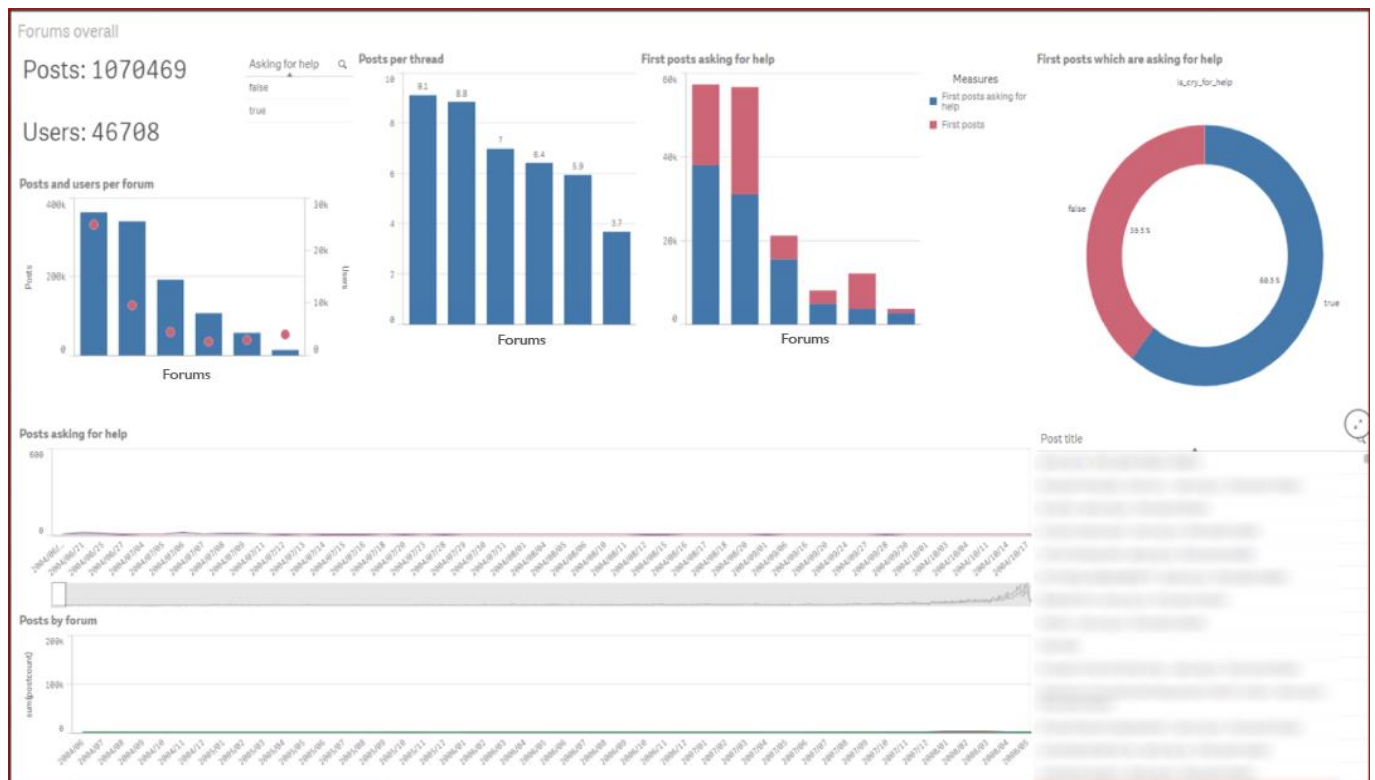
Researchers also used these prototype dashboards to identify additional fields needed for analysis, such as the position of a given post in a thread. The needs identified during this iterative process were addressed by modifying both the web scraper and the Python script used to process the scraped data.

After scraping and processing the data, each post contained the following metadata, which was used to build the final dashboard:

- An anonymised ID of the user who created the post
- The post's content
- The post's creation time
- The title of the thread a post belonged to
- The position of a post within a thread, and the number of posts within that thread.
- Which forum the post was from
- The original URL of the post

A final dashboard, pictured above, was used during the workshop held at The King's Fund with various health and data professionals. This dashboard included:

- Volume counts for unique users and posts in the dataset overall
- The same counts broken down for each forum
- Time series graphs showing post volume over the collected period, for each forum
- Various graphs showing the proportion of first posts classified as asking for help
- Overall user and post volumes
- A list of posts, with their titles and positions within the thread, with those asking for help highlighted
- A pane showing further details for single posts, when they're selected, including the entire text of the post
- A table showing the most frequently mentioned organisations within posts

OpeNER

In order to further enrich the data, we also a Named Entity Recognition (NER) Service, named OpeNER. This added three new fields to the data:

- Terms likely to refer to people
- Terms likely to refer to organisations
- Terms likely to refer to places

This NER Service uses a form of Conditional Random Field Classifier (CRFC). A CRFC for NER is trained on a large number of example named entities, and the contexts in which they arise (i.e. documents). It learns by identifying the consistent patterns which distinguish those named entities and, once trained, it can recognise named entities in documents by looking for the patterns learned during training.

On analysing the results of these, however, we found the terms suggested as people and places were of little use, and often inaccurately labelled. The organisation field, however, did produce some interesting results, and this was included in the final dashboard. Poor performance in this case can be explained by the fact that the NER model we used was trained on a set of input documents which were not representative of the forum posts seen in this project. This would mean that the patterns learned by the model were not characteristic of those named entities in forum posts and was therefore not well equipped to recognise them.

The following table shows a breakdown by site of relevant terms returned

by this process, in order of each term's prevalence within the collection.

Table 6

Patient advice forum	Mental health forum 1	Mental health forum 2	Carer's forum	Mental health medication forum	Depression forum
NHS	NHS	NHS	NHS	FDA	FDA
CT	MH	Mental Health	DWP	Pfizer	Tapatak
PSA	MentalHealth	Sane	BBC	American Psychiatric Association	Food and Drug Administration
TRY	BBC	TRY	Department of Health	APA	APA
FDA	DWP	CMHT	Care	Eli Lilly	Eli Lilly
DWP	ESA	DWP	ATOS	DSM	General Hospital Psychiatry
NICE	CMHT	BBC	Mental Health	GSK	University of Michigan Medical School
Apple	KS	ATOS	Care home	Department of Psychiatry	TRY
European Academy	Mental Health Forum	NHSDirect	RoyalTrust	Merck	Pfizer
MTX	FBI	MIND	National Autistic society	Ashton	NHS

Classifiers

NLP classifiers were built to determine whether our approach could offer insight into the *types* of conversations that were taking place on online health forums. Determining how to categorise data is an inexact science - previous research into forum thread types has identified a wide array of types.[17] Based on both the data available and the result of the workshop, we developed three lines of investigation, each aiming to test a different form of research.

For each of these, we constructed a separate architecture within Method52, often involving NLP classifiers. Each classifier within this project

was built using Method52's web-based user interface to proceed through the following stages.

Stage 1: Definition of categories

The formal criteria explaining how posts should be annotated is developed. Practically, this means that a small number of categories – between two and five – are defined. These will be the categories that the classifier will try to place each (and every) post within. The exact definition of the categories develops throughout the early interaction of the data. These categories are often not arrived at a priori, but rather iteratively, informed by the researcher's interaction with the data – the researcher's idea of what comprises a category will often be challenged by the actual data itself, causing a redefinition of that category. This process ensures that the categories reflect the evidence, rather than the preconceptions or expectations of the analyst. This inductive, data-driven approach to developing themes is consistent with a well-known sociological method called 'grounded theory'.

Stage 2: Creation of a gold-standard test dataset

This phase provides a source of truth against which the classifier's performance is tested. A number of posts (usually 100, but more are selected if the dataset is very large) are randomly selected to form a gold standard test set. These are manually coded into the categories defined during stage 1. The posts comprising this gold standard are then removed from the main dataset, and are not used to train the classifier. There are three outcomes of this test. Each measures the ability of the classifier to make the same decisions as a human in a different way.

- *Recall*: The proportion of data items in that class that are classified as being in that class.

- *Precision:* The proportion of data items that are identified as being in that class that are actually in that class.

- *Overall – F Score:* The 'overall' score combines measures of precision and recall, to create one, overall measurement of performance for the classifier.

All classifiers are a trade-off between recall and precision. Classifiers with a high recall score tend to be less precise, and vice versa. Scores for the classifiers built during this project are presented below.

Stage 3: Training

This phase describes the process wherein training data is introduced into the statistical model, called 'markup'. Through a process called 'active learning', each unlabelled post in the dataset is assessed by the classifier for the level of confidence it has that the post is in the correct category. The classifier selects the posts with the lowest confidence score, and these are presented to the human analyst via the user interface of Method52. The analyst reads each post, and decides which of the pre-assigned categories (see Phase 1) that it should belong to. A small group of these (usually around 10) are submitted as training data, and the NLP model is recalculated. The NLP algorithm then looks for statistical correlations between the language used and the meaning expressed to arrive at a series of rules-based criteria, and presents the researcher with a new set of posts which, under the recalculated model, it has low levels of confidence for.

Stage 4: Performance review and modification

The updated classifier is then used to classify each post within the gold standard test set. The decisions made by the classifier are compared with the decisions made (in Phase 2) by the human analyst. On the basis of this comparison, classifier performance statistics – 'recall', 'precision', and 'overall' (see 'assessment of classifiers', below) - are created and appraised by a human analyst.

Stage 5: Retraining

Phase 3 and 4 are iterated until classifier performance ceases to increase. This state is called 'plateau', and, when reached, is considered the practical optimum performance that a classifier can reasonably reach.

Stage 6: Processing

When the classifier performance has plateaued, the NLP model is used to process all the remaining posts in the dataset into the categories defined during Phase 1, using rules inferred from data the algorithm has been trained on. Processing creates a series of new data sets – one for each category of meaning – each containing the posts considered by the model to most likely fall within that category.

Classifier scores - Investigation 1 ('cries for help')

Classifier 1 - 'Cries for help'

Table 7

Label	Precision	Recall	F-Score	Accuracy	Coded
needs-help	0.538	0.724	0.618		47
Other	0.778	0.609	0.683		61
				0.653	

Classifier 2: 'Sought help in the past'

Table 8

Label	Precision	Recall	F-Score	Accuracy	Coded
has-sought	0.59	0.783	0.673		36
not-sought	0.565	0.371	0.448		31
other	0.438	0.368	0.4		37
				0.56	

Classifier scores - Investigation 2 (CBT)

Classifier 1: 'Mentions of a personal experience of CBT'

Table 9

Label	Precision	Recall	F-Score	Accuracy	Coded	Prior Multiplier
have-had	0.652	0.804	0.720		61	15
Other	0.645	0.455	0.533		65	1
				0.65		

The goal in building this classifier was to enable researchers to identify a group of participants who were likely to have had CBT. To do this effectively, the classifier needed to obtain a high overall score for the label 'have-had', above, which was achieved to a reasonable extent.

While building this classifier, a number of points stood out. In order to reduce ambiguity, and lower the prevalence of edge cases, we required posts labelled 'have-had' to explicitly mention therapy or experience. There were, however, a number of cases where experience with CBT was strongly implied – for example, by a user dispensing detailed advice or opinion of CBT – but could not be explicitly confirmed. While these posts were classed as 'other', they often shared terms indicative of the 'has-sought' category, which may have made the classifier less able to distinguish between the two.

ANNEX 2: FORUMS STUDIED

Patient advice forum

A collection of health-related forum boards, allowing users to discuss and check symptoms for a variety of conditions. This forum was chosen, in part, due to its large size and focus on dispensing health advice. We concentrated our collection on the site's sizable section on mental health.

Mental health forum 1

Smaller in size than the patient advice forum above, this forum is a general space for discussing mental health issues, and maintains sub-forums for a variety of issues (depression, schizophrenia and self-harm, for example)

Mental health forum 2

This forum is run by a UK based mental health charity, whose website provides a range of resources for those affected by mental health issues. Their support forum is one of these resources, and focuses on broad issues rather than specific conditions.

Depression forum

This is a small forum focussed on depression and depressive disorders.

Carer's forum

This forum is run by a charity which supports those who care for people with long term conditions in the UK. They run a highly active forum which allows users to discuss specific disabilities and conditions, as well as caring in general.

In addition to the five forums found through search engines, we also searched through the links posted in the forums themselves, in a process described below. This allowed us to identify a further forum which was added to the collection:

Medication forum

This site connects users who have experiences with medication commonly prescribed to treat mental health issues, and contains sections on self-care, dealing with the symptoms of withdrawal, and sharing success stories.

NOTES

[1] http://www.nhsconfed.org/~/media/Confederation/Files/Publications/Documents/the-futures-digital.pdf

[2] http://blogs.scientificamerican.com/guest-blog/accuracy-of-medical-information-on-the-Internet/

[3] https://dayinthelifemh.org.uk/

[4] https://www.nesta.org.uk/sites/default/files/peer_support_what_is_it_and_does_it_work.pdf

[5] http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(15)60394-4.pdf

[6] Leaman R. et al, 'Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts to Health-Related Social Networks', 2010: http://dl.acm.org/citation.cfm?id=1869976

[7] Hyun S. et al, 'Exploring the Ability of Natural Language Processing to Extract Data from Nursing Narratives', *Comput Inform Nurs,* 2009. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4415266/

[8] All quotes have been masked by researchers. This means they have been changed slightly to prevent the original post being found, but the meaning has not been changed.
[9] http://www.kingsfund.org.uk/publications/long-term-conditions-and-mental-health

[10] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969247/

[11] Bartlett et al (2014) *Vox Digitas*, Demos

[12] Davey et al (2015), 'e-Psychonauts: Conducting research in online drug forum communities' http://www.tandfonline.com/doi/abs/10.3109/09638237.2012.682265#.VbtWevlViko

[13] http://www.nhsconfed.org/~/media/Confederation/Files/Publications/Documents/the-futures-digital.pdf

[14] Method52 is a software suite developed by the project team over the last 18 months. It is based on an open source project called DUALIST - Settles, B. (2011) Closing the Loop: Fast, Interactive Semi-Supervised Annotation With Queries on Features and Instances. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1467-1478. It enables non-technical analysts to build machine-learning classifiers. The most important feature of is the speed wherein accurate classifiers can be built. Classically, a natural language processing (NLP) algorithm would require roughly at least 10,000 examples of 'marked-up' examples to achieve 70 per cent of accuracy. This is both expensive, and takes days to complete. However, DUALIST innovatively uses 'active learning', an application of information theory that can identify pieces of text that the NLP algorithm would learn most from. This radically reduces the number of marked-up examples from 10,000 to a few hundred. Overall, in allowing social scientists to build and evaluate classifiers quickly, and therefore to engage directly with big social media datasets, Method52 makes possible the methodology used in this project.

[15] Simon Wibberley, David Weir and Jeremy Reffin (2014) Method52 for Mining Insight from Social Media Datasets. In Proceedings of the 25th International Conference on Computational Linguistics. System Demonstration. (COLING 2014).

[16] See http://nutch.apache.org/ and http://lucene.apache.org/solr/

[17] Finn, 'An exploration of helping processes in an online self-help group focusing on issues of disability', *Health & Social Work*, 1999.

The Centre for the Analysis of Social Media is a collaboration between Demos and the University of Sussex. The Centre combines automated data extraction and sentiment analysis with social science statistics, analysis and ethics, to produce insightful and robust policy research.

Josh Smith is a Researcher in Demos' Centre for the Analysis of Social Media. He has a background in software development and philosophy, with a particular interest in the application of natural language processing to social scientific research.

Jamie Bartlett is the Director of the Centre for the Analysis of Social Media at Demos. Prior to working for Demos, Jamie was a research associate at the international humanitarian agency Islamic Relief and conducted field research in Pakistan and Bangladesh. Jamie holds Master's Degrees from the London School of Economics and the University of Oxford.

David Buck is Senior Fellow, Public Health and Inequalities, at The King's Fund. Before joining the Fund, David worked at the Department of Health as deputy director for health inequalities. He managed the Labour government's PSA target on health inequalities and the independent Marmot Review of inequalities in health. He has also worked at Guy's Hospital, King's College London and the Centre for Health Economics in York.

Matthew Honeyman is a researcher in The King's Fund's Policy team. He has a special interest in the relationship between health care, public policy and digital technology, and writes about how new technology can be deployed in the health system. Before joining the Fund, Matthew worked as a researcher and co-ordinator at the Innovation Unit, a social enterprise that works with public services to reshape the services they deliver.