

Smart IoT and Soft AI

*R Milton**, *D Hay†*, *S Gray‡*, *B Buyuklieva**, *A Hudson-Smith◊*

**richard.milton@ucl.ac.uk, †d.hay@ucl.ac.uk, ‡steven.gray@ucl.ac.uk,*

**boyana.buyuklieva.14@ucl.ac.uk, ◊a.hudson-smith@ucl.ac.uk,*

all authors from The Bartlett Centre for Advanced Spatial Analysis, University College London, England

Keywords: chatbots, language, AI, disclosure, gnomes

Abstract

Soft artificial intelligence (AI) is defined as non-sentient AI designed to perform close to human level in one specific domain. This is in contrast to “Artificial General Intelligence” (AGI) which solves the problem for human level intelligence across all domains. Soft AI is a reality now in the new generation of smart Internet of Things devices like Amazon’s Alexa, Apple’s Siri or Microsoft’s Cortana, giving rise to concerns about privacy and how the technology is being used. This research is based around an experiment in “AI as a service” where fifteen chatbot agents using Google’s “Dialogflow” are deployed around the Queen Elizabeth Olympic Park in London for the general public to interact with. The physical devices are 3D printed representations of creatures living in the park, designed to fit with the park’s biodiversity remit. Park visitors interact with the creatures via their mobile phones, engaging in a conversation where the creature offers to tell them a memory in exchange for one of their own, while warning them that anything they say might be repeated to others.

The scope of the work presented here is as follows. After explaining the details of the deployment and three month study, the conversational data collected from visitors is then analysed. Following a review of the current literature, techniques for working with the unstructured natural language data are developed, leading to recommendations for the design of future conversational “chatbot” agents. The results show distinct patterns of conversation, from simple and direct “verb plus noun” commands to complex sentence structure. How users interact with the agents, given that they are conversing with a mechanism, is discussed and contrasted with the memories that they have agreed to share.

The conclusion drawn from this work is that, while the current generation of devices only listen for commands from users, there is a danger that smart IoT devices in the future can be used as active information probes unless properly understood and regulated. We finish with observations on privacy and security based on our experiences here.



Fig. 1: Fifteen of these plinths displaying 3d printed creatures were deployed in the Olympic park for 3 months. This is Loki the gnome, which is located outside the View-Tube cafe area near the cycle-hire. The plinth contains a Bluetooth Beacon which alerts people to his presence via ‘Physical Web’ notifications on their mobile phones.

1 Introduction

This research is based around the deployment of fifteen “Geolocated Natural language Objects for Memory Evocation”, or GNOMEs for short, which are currently located in the Queen Elizabeth Olympic Park. Visitors are prompted to have a conversation with the gnomes via their mobile phones, during which time the gnome will ask them to leave a memory about the park. This work is motivated by the recent popularity in natural language user interfaces, as demonstrated by Amazon’s Alexa and a myriad of similar devices. We use the term ‘command and control’ to describe the current speech interfaces, which are typified by interactions of the form, “do I need an umbrella today?” or “turn the lights on”. By contrast, a full conversational interface raises a number of serious security and privacy questions. Psychology and social engineering plays a part here, as, once the computer is able to form its own questions to ask, then it ceases to be a passive user interface and instead becomes an active probe for information. The success of IBM Watson’s ‘DeepQA’ algorithms in beating human players at the game of ‘Jeopardy’ gives a hint as to what might be possible in the future [1]. In a game with no rules and with a connection to the Internet, an intelligence that can link data and probe for answers is a dangerous proposition.

While the chatbots used in this experiment lack the level of sophistication just described, there is an interesting duality between the data analysis and the ability of the chatbot to hold a conversation with a human. The gnomes deployed in the park use Google’s ‘Dialogflow’ API for the chat interface. This is a deep learning system which has been trained on the huge corpus of natural language text available to an Internet giant like Google. Where the data analysis of the memories left by the park users would historically have been of a qualitative form, the natural language processing techniques behind how contemporary systems Dialogflow work now allow for automated and quantitative analysis. The ability of the computer to hold a conversation with a human requires the computer to have some knowledge about how language is used, knowledge and algorithms which are then used in our analysis of the conversations and memories left by users, which then guides us as to how to improve the conversational agent. The reinforcement learning works towards continual system improvements, up to an as yet unknown cognitive limit.

The analysis is necessarily circular, with understanding of the data leading to better conversational algorithms. One of the failings discovered in the Dialogflow system is its inability to detect non-English names, while detection of the memories expressed by users in their conversation text is also a hard computational problem. Taking this a step further, our aim is to be able to contextualise the content of the conversation and quantify the information it gives us about the real world. The work that follows is an analysis of a ‘Dialogflow’ deployment in the real world, which we use as a seed experiment to analyse and improve on for the next generation deployment.



Fig. 2: Four IoT devices: models of a gnome, bat, otter and honey bee. Fifteen were 3D printed and deployed around the Queen Elizabeth Olympic Park with Eddy-stone Beacons for the general public to interact with. The models were painted in bright colours by local children from a community group between the ages of 5 and 15.

2 Deployment

Fifteen 3D printed creatures, comprising 7 gnomes, 3 otters, 3 bats and 2 honey bees are sited in 3 geographic clusters around the Queen Elizabeth Olympic Park (QEOP). Figure 2 shows the four basic models, which were then painted in bright colours by children from a local community group between the ages of 5 and 15. The deployment runs from 1st September to 14th December 2017, allowing us to capture 3.5 months of data from park users. The image in figure 3 shows the layout of the 560-acres-wide park with the locations of the venues and creatures, while table 1 shows the creature names and a summary of the data collected. The map is available to users on their mobile phones and has been designed to look like a computer game. Locations vary between exposed hilltop, entrances to venues, West Ham thoroughfare, lakeside and residential. Location choice is dictated by the creature’s species, with the otter models sited near to the ‘Timber Lodge’ and ‘Aquatic Centre’ where real otters are expected, the bats are sited along waterways and co-located with ultrasonic bat detector boxes, while the two honey bees and seven gnomes are distributed throughout the rest of the park.

The chatbot interface was developed using Google’s ‘Dialogflow’ system¹, using a website to host the park map and chatbot interface so that the interaction between the user and the creature happens via their mobile phone and the ‘tales of

¹ For Dialogflow, see: <https://dialogflow.com>

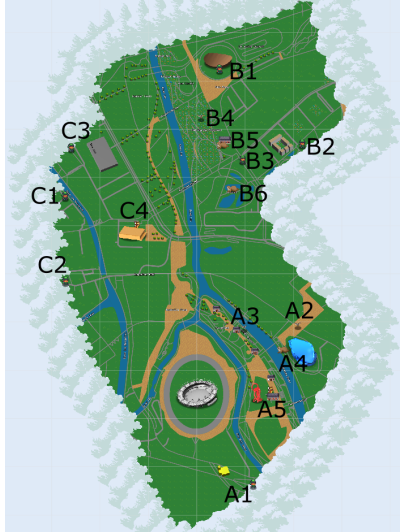


Fig. 3: This image shows the map of the park used on the mobile website. The distance from the top to the bottom of this map is 2KM. The alphanumeric codes denote the creature number, which links with table 1. Image: <http://talesofthepark.com>

the park’ website. In addition to this, the GPS position from the phone is used to limit conversations with creatures to within a 200 metre radius of their physical location. As people walk through the park, if they have ‘physical web’ and bluetooth enabled on their phones, then they will receive notifications as they approach the vicinity of a creature, prompting an interaction. The link to the website, <http://talesofthepark.com>, to initiate a conversation with the creature is included in the ‘Eddystone URL’ bluetooth packet, so the user only needs to click on the link. Instructions for how to chat with the creatures via a direct link are also printed on the plaque on the front of the plinth.

3 Background

When Joseph Weizenbaum published the ELIZA program in 1966, it applied decomposition rules to natural language inputs typed by the user to elicit a conversation [2]. His statement that, “the ELIZA script contains *no* information about the real world is not entirely true”, is a reference to the fact that ELIZA’s only knowledge about the outside world is through the encoding of its language transformation rules and nothing else. To compare this to a modern ‘Deep Learning’ architecture, we can look to the ‘Sequence to Sequence’ (SEQ2SEQ) approach of Sutskever on machine translation of natural language [3]. Here, the task of translating text from one human language into another is accomplished by training the system on a large corpus of text. Each sentence is converted to a high dimensional representation of a ‘concept’ so that two similar concepts with different sentences are mapped close to one another. This relies on the technique of word embedding [4], as

Code	Name	Lines	MemYes	MemNo	Others
A5	Rosie	646	28	13	9
B6	Denchu	496	19	5	15
A2	Moonlight	366	22	1	5
A3	Shadow Blade	276	12	2	5
B2	Jetpack Gnomey	270	10	2	3
A4	Zack	219	6	2	3
B5	Khadija	211	6	1	4
C3	Parker	199	8	4	1
C4	Beehigh	185	4	4	1
C1	Wombat	157	2	1	8
B1	Yusuf	121	4	0	3
A1	Loki	89	3	0	7
B3	Super Gnome	76	4	1	2
B4	Goku	47	1	0	2
C2	Gnomeo	12	0	0	1

Table 1: The raw interaction data for all the creatures collected over the two week period. The ‘Lines’ column is the total number of conversation lines, ‘MemYes’ is the number of times a user agreed to leave a memory, while ‘MemNo’ is the number of refusals. The ‘Others’ column is the number of times this creature was visited by a user who had previously visited one of the other creatures.

used in natural language processing, for example Mikolov [5] compares ‘Continuous Bag of Words’ (CBOW) and n-Gram similarity measures with word vector representations from 50 to 1000 dimensions to measure accuracy on different translation tasks. This follows on from his PhD thesis on neural network language models in 2012 which demonstrated their effectiveness in language modelling [6].

The motivation behind this work was also to improve on the current ‘Dialogflow’ chatbot, based on our analysis of breakdowns in the conversation. Repetition of the request by the user is a clear indicator of when this breakdown occurs and our agent is responding in the wrong way. Through analysis of the user responses, we can gauge their information content, which serves to improve the chatbot’s performance. This leads to the main result from this work, which is to show the potential privacy and security concerns which arise from an AI agent that can engage in a two way conversation. Taking a simplistic view of chatbot technology, the mechanism behind how ‘Dialogflow’ works matches text to a context pattern supplied by the programmer, which then triggers an action by the program. For example, given the text, “*can you tell me where the next creature is please*”, or even, “*next creature*”, the chatbot program enters a state appropriate for its response. However, where users leave their memories about the park, this state is much harder to detect. In the following section we define the language model used for the analysis, before applying it to the task of testing the data we collected for potentially disclosive facts.

4 Language Models

In order to create a language model, all the text collected for the three months that the creatures were deployed in the park was cleaned and converted to a sequence of integer word identifiers based on word frequency. This resulted in a word dictionary containing the 1,066 unique words which compose the 36,347 word corpus² used for training. Word vectors are learnt from training the language model to predict a word in the sequence, using a subset of the surrounding words as input.³ Given the context that words are used in, the model builds word vector embeddings where words with similar meanings are close together. For the purposes of this experiment, a word embedding vector size of 512 was used. A common technique with a large text corpus is to order the words by frequency and drop any below a threshold on the basis that they are either misspellings or unimportant. Here, though, the group of lower frequency words corresponds to the data entered by the users due to the fact that the ‘Dialogflow’ intents and actions scripts repeat the same sentences. These low frequency words correspond to the memories and other information that we are trying to extract from the text, including visitor names.

Using this type of word embedding, direct classification of the sentences in our conversational data is possible. We investigated the ‘gensim’ library’s implementation of the ‘word2vec’ algorithm [7], using vectors of 100 and 512 dimensions for comparison. In addition to this, a Tensorflow implementation [8] using a recurrent neural network (RNN) was also tested to see if it led to improved classification by overcoming the fixed length input limitations of the CBOW and skip-gram techniques while also preserving word ordering⁴. This used ‘long short term memory’ (LSTM) cells [10] [11], which are trained in batch mode over the input data with a final softmax layer used to flatten the output states from the LSTM cells and classify the output into a prediction of the most probable next word. A cell size of 512 was used with a 2 layer deep RNN, time step size of 35 and dropout 0.5. Using this architecture, a sentence context state is stored using $512 \times 2 \times 2 \times 35 = 71,680$ floating point values. The results from the word embedding comparison showed that the park corpus is too small to show complex relationships between words, but is nonetheless useful due to the domain specific nature of the conversations about the park. The word, ‘gnome’, rarely appears in American English, while the park corpus is able to link the ‘ArcelorMittal’ venue with its common misspelling, ‘ArcellorMittal’ based on cosine distance. Of particular interest in detecting personal data are pronouns like, ‘I’, ‘he’, ‘she’, ‘they’, where the following statement is usually a fact about something that happened in the park. When detecting conversations about places,

² By way of comparison, Shakespeare’s Hamlet is 30,557 words.

³ It is usual to build this type of language model on a large corpus of text, for example the Penn Tree Bank (PTB), Brown corpus, Text8 or Wikipedia dump, where the vocabulary size could number 70,000 words. Variations on the use of a large corpus in the context of the park data are currently being investigated.

⁴ See [9] for a related technique using paragraph vectors which also rejects the bag of words approach.

though, the park corpus is able to link the ‘Timber Lodge’, ‘Cafe’ and ‘coffee’ in a way that is not otherwise possible. The conclusion drawn from the word vector analysis is that local domain specific information is essential, requiring a fusion between the small scale and large scale text analysis methods, for example following Taddy’s methodology on sub-corpus partitions [12].

5 Data Analysis

During the data collection period, a total of 4,186 lines of conversation were recorded in 186 distinct chains of conversation. This can be broken down further into 3,370 lines from the creatures and 816 from the users. Rosie the bee, located at the foot of the ArcelorMittal Orbit attraction was the most talked to creature with 646 conversation lines, 31 user interactions and 28 memories left. The data is summarised in table 1. Interaction rates for the creatures are very low compared to the expected footfall for the locations. From the park’s attendance figures for 2016, there were 101,633 visitors to the ArcelorMittal Orbit that year. A rough calculation puts the interaction rate at 0.1% of the orbit’s visitors talking to Rosie.

By way of example, a typical conversation with one of the creatures is as follows:

On 12 September 2017 at 15:13 Anne talking to Denchu the otter:

...

Denchu: If you like I can tell you more about the River Lea and the Waterways

Denchu: Or I can direct you to the next Creature

Anne: Tell me more please

Denchu: The Park’s waterways and wetlands have been identified as a great habitat for Otters, and one day the Park hopes to be home to some!

Denchu: What other species would you like to see in the Park?

Denchu: If you’d like to tell me something about this, answer ‘yes’

Anne: Yes

Denchu: Great! What would you like to tell me?

Anne: I’d like to see sharks here

Denchu: Really, well that’s really interesting. I’ll remember that.

Denchu: In return, here’s a fact about the River Lea and the Waterways

...

On 27 October 2017 at 13:33 Lisa talking to Denchu the otter:

...

Denchu: Great! What would you like to tell me?

Lisa: We fed the ducks today

Denchu: Really, well that’s fascinating. I’ll remember that.

Denchu: In return, here’s a memory that Anne left with me earlier

Denchu: “I’d like to see sharks here”

Denchu: Would you know anything about that?

Denchu: Answer ‘yes’ if you’d like to tell me something

Lisa: I Don’t think there are sharks here but a duck bit me

...

An analysis of the user names recorded reveals the problems in the Dialogflow scripts: “Hello Rosie I’m a teacher”, “Hi Parker, I’m Jack” and “Sorry where is the next one” appear amongst Anne, Lisa, Connie, Katie, Dave and others. This occurs where a breakdown happens in the initial part of the conversation script that deals with the names, leading to the question and answer sequence becoming out of step. When given anything more than a simple response to the name question, our ‘Dialogflow’ script parses the name incorrectly. What is required here is the ability to split a response into separate clauses and make an inference about whether a word constitutes a name or not.

By taking only the responses from the visitors and ordering them by sentence length, it immediately becomes apparent that the memories they have chosen to leave are the longest parts of the conversations. At 59 word tokens, the longest memory left is as follows: “I’ve been rollerblading around the Olympic Park today, it’s a wonderful place for it. Everywhere is very accessible on skates, and there’s so much to explore. You can stick with the crowds, or find quiet corners where it feels like you’re the only person in the world. It’s great to watch landmarks like The Orbit as you glide by.”

The original hypothesis was to detect whether the sentence structure changes as a result of the realisation by the visitors that they are talking to a mechanism. By plotting how the length of sentences varies between the beginning and the end of a conversation, figure 4 shows a strong bias towards sentence length increasing. Where the opposite is true can then be used as an indicator of where the ‘Dialogflow’ chatbot is failing. We referred to this as the ‘command and control’ hypothesis, but it is closely related to the information content of the text as just described in the length of the memories left. While analysis of large-scale texts, for example the ‘Text8’ corpus previously mentioned, discard words that appear below a threshold number, these are the words forming the highest information parts of our small-scale park text corpus. We are also interested in memories that could lead to unintended information disclosure, so, for example, “I got married in the Olympic park [date redacted], Fish Island and Formans after the Orbit”, is enough information to start looking for wedding photographs on the Internet. The design of a privacy violating chatbot then becomes a two stage search algorithm, where the computer first determines what the person has said before looking for a response to extract further information. This is looking at how to minimise the search space further by asking a follow-up question. When ‘Dave’ says about the Arcelor-Mittal tower, “The view is good, and you can see my house from the top, but the ticket prices are much too expensive.”, a strategic follow-up question could further narrow down where he lives. When this is seen as a case of using intersecting partial information to narrow down a search space, an observation that we can make from the data collected so far is that either one of *spatial* or *temporal* information is valuable, with a third class of *semantic* information that can be used to derive one of

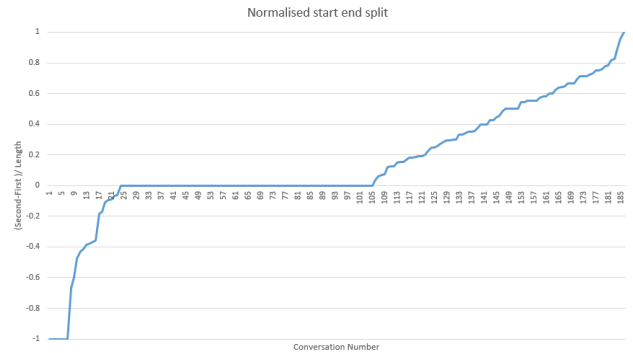


Fig. 4: Split between number of words at the beginning (W_b) and end (W_e) of conversations. This is defined as $(W_e - W_b)/W$, where W is the total number of words. The predominantly positive ratio suggests more complex text at the end of conversations where people leave memories, but with a small cluster of language degeneration at the negative end which needs further investigation. This shows more complex language than “verb + noun”.

the other two. For semantic, we include things like, ‘Danny’⁵ saying, “I’ve come to the park to do fieldwork for geography gcse”. This gives the chatbot an age and a name to look up in any public information on local schools in the area. Here, though, the name can have a large impact on how far the search space can be narrowed down. Using the Office for National Statistics baby names data for 1996 [13], we find that the probability of a boy named ‘Aaron’ is 7.88×10^{-3} , compared to 6.28×10^{-5} for the alternative spelling ‘Aarron’, while there were 325 babies named ‘Danny’ in 2002, who would now be doing their GCSEs. This official dataset on baby names provides us with the ability to match any UK registered names in the visitors’ responses, providing us with an indicator variable and an ability to quantify the disclosiveness.

6 Conclusion

Even in the small sample of 186 park conversations collected here, we can detect one privacy concern that could be exposed with a combined spatial and temporal search, “we got married at the orbit on [date]”, and another spatial statement of the form, “you can see my house from here”. While the first is detected by a filter trained to identify months of the year, the second depends on a deeper understanding of language. Figure 5 shows three of the indicator traces for the park data, based on a combined ‘Text8’ and ‘park’ word corpus.

While this simple chatbot ‘game’ only serves as a lure for data, our analysis shows how the context matching used in the chatbot’s state transitions could be augmented with information from the Internet to steer the AI’s response towards further in-

⁵ The name has been changed to preserve anonymity.

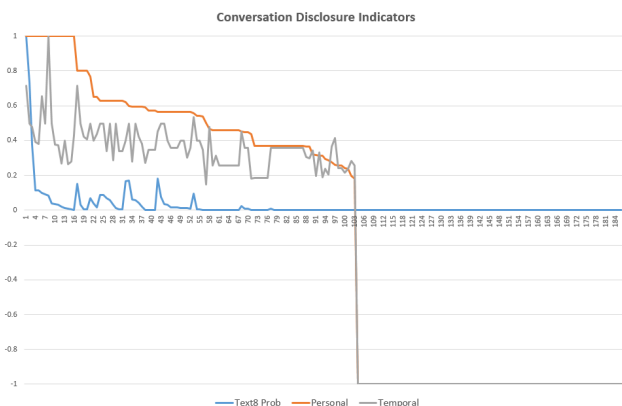


Fig. 5: Conversation disclosure indicators, sorted by personal disclosure factor, ‘I’, ‘he’, ‘she’ etc. Indicators are normalised to ± 1 for comparison. Positive means strong association. The -1 temporal value from conversation 103 onwards indicates too little text to obtain any data. Two further traces based on word length and entropy have been omitted for clarity.

formation disclosure. This paper shows that advanced natural language understanding is not required to achieve this, only the ability to pick out a key part of the response to prompt the visitor for further information, much like the original Eliza program did in 1966.

As further advancements are made in natural language processing, we find ourselves in the area of ‘content overhang’, where large-scale unstructured data is available on the Internet to any algorithm that is able to use it. As highlighted earlier, if the requirement is to build a better chatbot, then that requires improved language capabilities, which leads to improved handling of unstructured text data from the Internet. The algorithmic potential could be there for a large-scale ‘take-off’ in capability. This is not command and control, this is poking for data.

7 Acknowledgments

The authors would like to thank EPSRC for funding this research through the Privacy, Ethics, Trust, Reliability, Acceptability and Security (PETRAS) grant (EP/N02334X/1), Cyber Security of the Internet of Things and IoT in the Park projects. Also, we would like to thank the London Legacy Development Corporation (LLDC) and Queen Elizabeth Olympic Park, without whose support this could not have happened. The OpenBeacons platform was funded through the Google Internet of Things Research Technology Awards Pilot study, who donated the Eddystone beacons used for this research. Finally, thank you to Michaela, Marcus, Muhammad, Loreal, Sierra, Mahjuza, Amina, Fotios, Michael, Elyas, Khadija, Dontae, Ijlaal, Hawa and Darnell (ages 5-15 years) who painted the models for us.

References

- [1] D. A. Ferrucci. Introduction to ‘this is watson’. *IBM Journal of Research and Development*, 56(3.4):1:1–1:15, May 2012.
- [2] Joseph Weizenbaum. Eliza - a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45, January 1966.
- [3] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. *CoRR*, abs/1409.3215, 2014.
- [4] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. *J. Mach. Learn. Res.*, 3:1137–1155, March 2003.
- [5] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.
- [6] Tomáš Mikolov. *Statistical language models based on neural networks*. PhD thesis, Brno University of Technology, 2012.
- [7] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. <http://is.muni.cz/publication/884893/en>.
- [8] Martín Abadi et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. <http://download.tensorflow.org/paper/whitepaper2015.pdf>, 2015. Accessed: 9 January 2018.
- [9] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. *CoRR*, abs/1405.4053, 2014.
- [10] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997.
- [11] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *CoRR*, abs/1409.2329, 2014.
- [12] Matt Taddy. Document classification by inversion of distributed language representations. *CoRR*, abs/1504.07295, 2015.
- [13] Office for National Statistics. Statistical bulletin: Baby names in england and wales: 2016. <https://www.ons.gov.uk/peoplepopulationandcommunity>. Accessed: 9 January 2018.