

Social Media Analytics in Museums: Extracting Expressions of Inspiration

Museums have a remit to inspire visitors. However, inspiration is a complex, subjective construct and analyses of inspiration are often laborious. Increased use of social media by museums and visitors may provide new opportunities to collect evidence of inspiration more efficiently. This research investigates the feasibility of a system based on knowledge patterns from FrameNet - a lexicon structured around models of typical experiences - to extract expressions of inspiration from social media.

The study balanced interpretation of inspiration by museum staff and computational processing of Twitter data. This balance was achieved by using prototype tools to change a museum's Information Systems in ways that both enabled the potential of new, social-media-based information sources to be assessed, and which caused the museum staff to reflect upon the nature of inspiration and its role in the relationships between the museum and its visitors. The prototype tools collected and helped analyse Twitter data related to two events. Working with museum experts, the value of finding expressions of inspiration in Tweets was explored and an evaluation using annotated content achieved an F-measure of 0.46, indicating that social media *may* have some potential as a source of valuable information for museums, though this depends heavily upon how annotation exercises are conducted. These findings are discussed along with the wider implications of the role of social media in museums.

Keywords: museum; inspiration; social media; analytics; semantic analysis; natural language processing

1 Introduction

Inspiring visitors is a core function of museums; the UK Museums Association defines museums as places that: 'enable people to explore collections for inspiration, learning and enjoyment.' (UK Museums Association 2014), and yet previous research undertaken by the authors (Gerrard et al, 2014) indicated that inspiration is a complex concept with many constituent parts, and as such it is not defined consistently by museum staff. Measures such as counting visitor numbers indicate the impact of museums, but without recognising when visitors have been inspired, such measures do not provide a full picture of museums' value (Holden, 2006).

The intention of the research was twofold: firstly, we aimed to evaluate the potential for social media data to provide a valuable source of information for museums. Secondly, however, in choosing the concept of ‘inspiration’ as the key term with which to try and retrieve information about visits, we also had an opportunity to work with museum staff to investigate this concept and its potential usefulness as a way to structure thinking about relationships with visitors. These intentions are summarised by the following research questions:

1. Did considering ‘inspiration’ help museum staff think about their relationships with visitors productively?
2. Did the concept of inspiration help museum staff work with information based upon social media data more effectively?
3. Could the concept of ‘inspiration’ be used as a ‘hook’ to retrieve usable, valuable information from social media data related to museum events?

The remainder of this paper is organised as follows: section 2 provides the theoretical background to inspiration in museums, retrieving information from social media, and artificially intelligent Information Retrieval systems. Section 3 presents the prototype social media analysis system and study, and section 4 covers the results of using the system both in terms of its success at retrieving information, and with regard to how museum workers interpreted the information it produced. Section 5 provides a discussion and covers the limitations of the research. The research conclusion is presented in section 6.

2 Theoretical Background

The research described here began with a definition of inspiration as: *an experience, or set of experiences, combining rational thoughts and emotions, resulting in the expression or enactment of fresh ideas*. It is important to recognise that this is not a comprehensive definition of inspiration, rather one that is very museum-specific. It results directly from our previous research, which consisted of a review of Museum Studies literature and discussions with museum practitioners [Authors, 2014]. The definition emphasizes experiencing authentic, tangible objects (Soren 2009) that reflect important aspects of history, art, science or culture (Selwood 2010). It also describes how emotions form an important part of our experiences (Connolly 2002) but that to be considered valuable in a museum context, inspiration must also incorporate rational ideas that help us relate to society (Barrett 2011).

There are also many similarities between this definition of inspiration and the definition of *engagement* in Brodie et al. (2011, p. 254), which also refers to experience, emotion, creativity and learning. Engagement is an often-used term both in relation to museums *and* social media. Brodie et al. describe the way in which engagement is facilitated by social networks, and describe how use of the term has increased since 2005 alongside the popularity of major social media platforms. Social media has also had a big impact in the museum sector, which was quick to highlight its potential not only for marketing museum activities, but also for providing channels through which to spread museum knowledge (Cameron 2008) and receive knowledge back from visitors

(Russo 2011).

Social media analytics are also used, increasingly, by museums to help evaluate the impact of their events (Villaespesa 2015, 2013; Langa 2014; Finnis et al 2011; Giaccardi 2011), but such evaluation is often undertaken by museum staff using off-the-shelf analytics tools, which tend to provide quantitative measures of ‘reach’ and ‘engagement’ that are similar to counting attendance figures. This research assessed the feasibility of using a computational model to automate the process of extracting expressions of inspiration from visitors’ social media, which may ultimately allow much larger data streams to be analysed and understood in near real-time.

2.1 *Detecting Subjective Expressions*

The research involved developing a prototype system to retrieve *potential expressions / indications* of inspiration. There is a significant difference between such indications and actual *evidence of inspiration*: the social media text in question can be analysed for hints that inspiration *may* have occurred, and in doing so indicate which visitors *may* have been inspired, but such a system would need to be used in tandem with other research approaches in order for *actual evidence* of inspiration to be collected (see Conclusions for more discussion regarding this point).

Automatically detecting subjective expressions in natural language and social media posts is well established. Within the computational research fields of Information retrieval (IR) and Natural Language Processing (NLP), opinion mining and sentiment analysis have received significant attention (Pang and Lee, 2008; Ravi and Ravi, 2015). These are fields related to the task of detecting expressions of inspiration from social media messages, in that, broadly speaking, opinion mining and sentiment analysis are concerned with the computational study of subjective states, such as sentiments, emotions, opinions and experiences expressed within text.

With a few exceptions (e.g. Chang et al., 2015; Bertola and Patti, 2016) there has been relatively little systematic research effort in exploring the application of these techniques to the museum or heritage sector. Nevertheless considering the large volume of social media messages publicly shared by museums and their visitors, arguably, computational methods are increasingly becoming relevant to analysing such user generated content (Fletcher and Lee, 2012). Numerous algorithms and techniques for detecting emotional expressions exist, including word dictionary and semantic model approaches (i.e. ontologies), with a recent review of state of the art available in Ravi and Ravi (2015). The models used to represent and analyse different psychological states and emotions vary from simple positive and negative sentiments (e.g. Thelwall et al. 2012), Ekman’s emotional dimensions – i.e. Anger, Disgust, Fear... – (e.g. Sykora et al. 2013), to models such as Russell’s emotional circumplex of affect (Russell 1980; Choudhury and Counts 2012). Inspiration in its own right is a distinct concept and although related, does not directly fall under prior streams of sentiment analysis work. The task of detecting expressions potentially indicative of inspiration has not been addressed separately in existing published work and this work presents a novel technique that relies on a semantic model (or ontology) called FrameNet, described in detail within section 2.2 and 3.2, which generally allows a richer representation of vocabularies than a dictionary based approach (Ravi and Ravi 2015).

Subjectivity related to the concept of inspiration also extended to the

interpretation of evidence regarding its potential occurrence; hence methods intended to assess the impact of information upon human thought and action were appropriate (Klein and Myers 1999). The most appropriate of these collaborative methods was Action Research (AR) in an Information Systems context, resulting in a need to collaborate closely with museum staff to assess their interpretations of potential expressions of inspiration retrieved from visitors' social media content.

2.2 *Interpreting Complex Experiences and Situations*

The lexical knowledge base FrameNet developed at UCLA Berkeley (Baker 2014) was chosen as a resource with which to build a model of inspiration because its lexicon is arranged around psychological structures (so called Frames), which are used to interpret potentially complex experiences and situations (Fillmore 1976). The theory of 'Frames' was explored by Erving Goffman (who also worked at Berkeley in the 1960s) as a way of describing: "...schemata of interpretation (Goffman, 1974)" of the events and situations that people experience. A related theory that also emerged from Berkeley in the latter part of the 20th century was Lakoff and Johnson's Conceptual Metaphor Theory, which notes that metaphor is not limited to creative writing and speaking, but is instead so ingrained in the ways we use language that it must also be a key part of how we think; that the metaphors: "... reflect how concepts are organised in our minds (Tay, 2014:53)" This implies a strong link between how people describe their experiences textually, and the ways in which they may be 'framing' those experiences subjectively, and it is this link that underpins FrameNet.

FrameNet provides a source of ready-made models of situations, defined as Knowledge Patterns (Clark et al. 2004), several of which are similar to the state of 'being inspired' as defined above. FrameNet's organization of lexical data around Knowledge Patterns based upon experiences distinguishes it from other lexical knowledge-bases such as ConceptNet (Havasi et al. 2007) and WordNet. The former of these alternative resources is a large database of simple 'common sense' relationships, while the latter is a hierarchical arrangement of general vocabulary. Neither resource has the same focus upon psychological structures and related specialised vocabulary as FrameNet. A method to apply the Knowledge Patterns in FrameNet to Natural Language Processing (NLP) tasks has also been explored by Nuzzolese et al. (2011). The FrameNet entities most referred to in this paper are:

- **Frames:** the core definitions of experiences and situations around which FrameNet's knowledge is structured. For example, the Frame *Becoming_aware* defines the experience of discovering or learning something.
- **Frame Elements (FEs):** the key working parts of each Frame that are used to define the processes at work in a given experience or situation. For example, *Becoming_aware* contains the FEs *Cognizer* (the individual gaining awareness) and *Phenomenon* (the thing they are becoming aware of).
- **Lexical Units (LUs):** the key terms, retrieved from annotated text, that relate text back to a particular Frame: for example, *Becoming_aware* has the LUs *discover*, *find*, *learn* etc.

Using FrameNet to attempt to retrieve potential evidence of inspiration from Tweets therefore involved linking Tweet text to Frames via their LUs, and then

analysing the nature of the relationship between Tweet and Frame in terms of the FEs that might be contained in the Tweet text. One disadvantage of using FrameNet for social media analysis was that FrameNet data is based upon annotations of complete sentences of edited well-formed text that are different from Tweets (Baker 2014). Hence this research also enabled investigation of whether the advantages of using well-structured, relevant Knowledge Patterns outweighed the disadvantages of using a lexicon compiled from non-social-media text.

3 Research Design

The research consisted of a mixed approach balancing collaborative interpretation of the concept of inspiration (conducted with staff from Derby Museums) and quantitative computational processing of data from Twitter. Derby Museums is the main museum organisation in the UK city of Derby, in the East Midlands region, approximately 130 miles north of London. Spread across three sites, their annual review covering the period between April 2014 to March 2015 states they had 97,558 visitors, 103,282 visits to their website and 12,119 likes and friends on social media (Derby Museums 2015). They also have a comprehensive social media strategy (Rippleffect 2015). Both the Head of Museums and the Social Media Coordinator from Derby Museums devoted their time to the interpretive research. Tweets were collected in relation to two museum events described below. The two research strands were brought together with a manual annotation task undertaken by the museum staff and first author upon a random subset of the collected Tweets, from which ‘precision’ and ‘recall’ measures were calculated to evaluate the effectiveness of finding potential expressions of inspiration by relating Tweets to Frames. ‘Precision’ concerns the ability of the system to *find* relevant content; it is calculated by measuring how many of the pieces of information the system retrieves from an annotated set of test data are deemed relevant. ‘Recall’ refers to the ability of the system *not to miss* relevant content, and is calculated by measuring how many pieces of relevant information the system *fails to retrieve* from the same annotated set (Baeza-Yates and Ribeiro-Neto, 1999).

Twitter was chosen as a data source for three reasons; (i) Twitter’s API provided easy access to sample data by searching for hashtags; (ii) the data was in the public domain (with the public nature of Twitter made clear without ‘legalese’ to its users at the point of sign-up); and (iii) the organisers encouraged visitors to Tweet about their experiences before, during and after attending the events. One of the key implications of using Twitter as a source of data is the underlying biases of the platform towards the demographics of its user base. Similarly, the nature of using hashtags as a method of data retrieval from Twitter results in a self-selecting user sample that is likely to be biased. Both of these issues are discussed in Tufekci (2014) and their implications for this research are covered in the discussion and conclusion, however a positive factor that emerged from using Twitter was the ease with which smartphones could be used to Tweet: this meant that a substantial proportion of the data was created during the two events, therefore potentially at the point where visitors may have actually felt inspired.

Tweet data was imported into a Neo4J Graph Database, a data management approach that is particularly suited for working with interrelated data in a rapid, prototypical fashion (Neo4J 2013). Neo4J enabled parts of Frames and the Tweets themselves to be stored as nodes in a directed graph structure, with edges between the nodes indicating the relationships between the Frames, their FEs, their LUs, and the Tweets that contained those LUs.

3.1 *Museum events*

The two events around which Twitter data was captured both took place in the autumn of 2014. The first was the Derby Mini Maker Faire: a two-day-long exhibition and fair in which ‘makers’ (designers, developers, crafts-people, hobbyists and learners, with interests in design, technology and engineering) exhibited their work. The second event was MuseoMix UK 2014, during which 90 participants from around the globe spent three days developing prototype museum exhibits using items from Derby Museums’ collection. Both events were open to members of the public as well as participants. The first author also attended both events, and participated in the second, to witness the reactions of participants and visitors first-hand. Tweets related to these events were collected using the Twitter API to search for the publicised event hashtags (#DMMF14 and #MMUK14), mentions of the formal event names, and mentions of Twitter users’ accounts directly linked to the events.

3.2 *Use of FrameNet and LU Linking*

After the events had occurred and the Tweet data had been collected, keywords derived from the terms in the initial working definition of inspiration and their synonyms were used to search the online version of FrameNet (<https://framenet.icsi.berkeley.edu/fndrupal/>) for relevant Frames using its search engine. Decisions regarding relevance were taken at a meeting between the lead author and the museum staff. As an example of a typical Frame, Figure 1 shows the data from the *Subjective_influence* Frame, returned by a search for the keyword ‘inspiration’. Tweets returned by searches for the key Lexical Units (LUs) of potentially relevant Frames were also reviewed at the meeting with museum staff, further indicating the degree to which FrameNet contained relevant knowledge. Frame data is provided by FrameNet in XML documents, enabling import into the Neo4J database. Relationships between the LUs and Tweets containing those LUs were created using word-bounded searches for the stem LUs with their various suffixes by utilizing Neo4J’s built in Regular Expression search function. For example, the derivations of the LU ‘inspire’ from the Frame *Subjective_influence* were found in Tweets and linked to the ‘inspire’ LU using the search term: *inspir(e/es/ed/ing)*. Linking Tweets to LUs meant that all Tweets containing at least one LU could be retrieved from the dataset, though Retweets were filtered out of the results to eliminate redundant information.

Speed description
Spelling and pronouncing
Sports jargon
Stage of progress
Standing by
State
State continue
State of entity
Statement
Stimulate emotion
Stimulus focus
Stinginess
Store
Storing
Strictness
Studying
Suasion
Subjective influence
Subjective temperature
Submitting documents
Subordinates and superiors
Subsisting
Substance
Substance by phase
Subversion
Success or failure
Successful action
Successfully communicate message
Sufficiency
Suicide attack
Suitability
Summarizing
Supply
Supporting
Surpassing
Surrendering
Surrendering possession
Surrounding

Subjective_influence Lexical Unit Index

Definition:

An **Agent**, **Situation** or **Entity** has an influence on a **Cognizer**. The influence may be general; or it may be manifested in the **Cognizer**'s engaging in an **Action** as a consequence of the influence; or the **Cognizer** may be influenced in how they carry out a **Behavior** that they are engaged in already. Alternatively, a **Product** may be specified whose production or design was influenced by the **Cognizer**'s experience of the **Situation** or **Entity**.

The mediation of the **Cognizer**'s psyche distinguishes this frame from the Objective_influence frame, where dependent events occur automatically given the appropriate kind of influencing force. In this frame, by contrast, a **Cognizer** may perceive an influence yet not respond to it in any way.

These works had a profound effect and **INFLUENCED** **her** in the creation of a successful series capturing this sun-drenched region.

I'm sure if I asked Mattel what **INSPIRED** this call I would be told it's an original design, and not supposed to represent any actual car.

FEs:

Core:

Action [act] An Action that the **Cognizer** decides to carry out as a consequence of the influence of a **Situation** or **Entity**.
Excludes: Behavior
What factors **INFLUENCED** Nixon to issue the doctrine?

Agent [Age] The **Agent** acts to influence a **Cognizer**.
Excludes: Situation

Behavior [beh] A **Behavior** that the **Cognizer** is engaged in independently but whose execution is subsequently influenced by the **Situation** or **Entity**.
Excludes: Product
Attention will also be paid to St. Paul , what **INFLUENCED** him in his thinking and what he wrote in his letters about sex, sexuality, men and women

Cognizer [cog] The person who is influenced by the **Situation** or **Entity**.
Semantic Type: Sentient
George Gilder would run with this idea later in writing the New Testament of supply-side economics

Figure 1 Subjective_influence Frame retrieved by searching FrameNet for 'inspiration'

3.3 Annotation Tasks and Evaluation

At this point, two annotation tasks took place. The first, conducted by the first author, annotated all 1165 original Tweets (i.e. not Retweets) that were linked to Frames. The FE structure of the Frame in question was also used to annotate parts of each Tweet, where possible. An initial assessment was also made of whether, in the opinion of the first author, each Tweet contained a relevant expression of inspiration. Because the first author's subjective opinion was key to making the decision regarding the relevance of each Tweet to the topic of inspiration, a subset of the collected Tweets was then re-assessed (by the first author once more) after a 14 month period had elapsed, in order to generate a measure of confidence in the reliability of the assessment (Krippendorff, 2004). The results of this 'intra-coder reliability measurement' are discussed in section 4.5.

The second annotation task involved a set of 400 original Tweets extracted from the dataset. The extraction was carried out manually to ensure that a high proportion of Tweets that had been linked to Frames were included, though all evidence of whether a Tweet was linked to an LU was hidden from the museum staff. They were then asked to annotate each Tweet with a yes / no answer to the closed question: 'does this Tweet indicate that someone may have been inspired?' As the busy museum staff were volunteering their time, long discussions to achieve a consensus regarding each Tweet were not practical, so instead the staff were asked to annotate the Tweets individually in four subsets of 100, with a short discussion about differences of opinion conducted between each set. The degree of consensus was recorded after each subset's annotation. Scores for precision and recall were then taken using an average of the two sets of opinions. For a further indication of how much staff-member bias was a factor, the first author conducted the same annotation task and his answers were compared (see Discussion and Limitations).

4 Results

4.1 Overview of the collected Tweet data

Table 1 contains information about the complete dataset of Tweets related to both events. The figures show the total number of Tweets collected via Twitter API searches, the number of ‘original’ Tweets (i.e. those not labelled a Retweet by Twitter), the number of *distinct* Tweets containing at least one LU from a relevant Frame, and the subset of those annotated ‘potentially relevant’ during the initial annotation task carried out by the first author.

The MuseoMix dataset is approximately 5.5 times larger than the Maker Faire set, due to the following factors; the event ran for a day longer and (according to the observations of the first author) the participants in MuseoMix were encouraged to Tweet about their experiences heavily by the event organisers. The proportions of Tweets linked to LUs from Frames deemed relevant to inspiration were similar across the two events, however (24% Maker Faire / 22% MuseoMix), and the proportions of Tweets linked to LUs that were noted ‘potentially containing expressions of inspiration’ were near identical (8% / 7%). These proportions are comparable to some prior work, for instance on emotion detection, where a larger set of basic emotions tended to occur in 12% of all the sampled tweets (Sykora et al. 2014).

Table 1. Overview of Tweet Data

Event	Maker Faire	Museo-Mix	Totals
All Tweets	1006	5652	6658
Original Tweets	469	2643	3112
Originals Linked to LUs	210	995	1165
Distinct Linked to LUs	111	588	699
‘Relevant’ Linked to LUs	38	198	236

4.2 Finding Frames Related to Inspiration

The Frame-finding exercise described in section 3.2 returned nine relevant Frames from FrameNet, listed in Table 2, which summarises how the Lexical Units (LUs) from the Frames mapped onto the Tweet data. The nine frames were categorised into three broader categories that related to the initial working definition of inspiration: (i) *cognitive* Frames related to having thoughts and ideas; (ii) *emotional* Frames related to experiencing feelings; and (iii) *creative* Frames related to producing and making things. These distinctions are evident in the descriptions of the Frames themselves – all but one of the Frames contains a ‘sentient’ Frame Element (FE). In cognitive Frames such as *Becoming_aware* and *Cogitation*, the sentient is called the *Cognizer*; in emotional Frames the sentient is the *Experiencer*; and in creative Frames the sentient is the *Creator*.

Frames may involve both Core Frame Elements and optional ones, for the sake of simplicity during this feasibility study, only Core ones were considered. The number

of FEs per frame is a good indicator of a Frame’s complexity: most of the Frames were quite simple (e.g. two FEs: the sentient and a phenomenon of some kind), but the *Subjective_influence* Frame (returned by a search for the term ‘inspiration’ itself) was much more complex, with a potential seven core FEs related to the sentient, the various types of stimulus that might influence them, and the various types of outcome. In contrast, *Subjective_influence* also has the fewest LUs related to it. The two emotional Frames had the greatest proportion of related LUs, though this was not reflected in the proportion of Tweets that were linked to them: only 19 of the 186 LUs in the *Emotion_directed* Frame were contained in the Tweets (10.2%), while 29 of 127 (22.8%) of *Experiencer_obj*’s LUs were found. Overall, the creative Frames had much broader overlap with vocabulary used in the Tweets (64% of LUs used in both *Creating* and *Intentionally_create*), but even they contained a large proportion of LUs that were unused. I.e., the Tweets contained a shallower vocabulary than FrameNet, supporting Baker’s assertion that FrameNet may not be entirely suitable for analysing social media text at present (2014).

Table 2. Summary of Frame, LU and Tweet relationships

Frame name	Category	# Core FEs	# Matched Lexical Units			# Original Tweets linked to frame					
						Maker Faire		MuseoMix		Both	
			Total	%	#Linked	Total	%	Total	%	Total	%
Becoming_aware	Cognitive	2	32	7%	16	21	10%	82	9%	103	9%
Cogitation	Cognitive	2	20	5%	7	6	3%	78	8%	84	7%
Coming_to_believe	Cognitive	4	21	5%	7	17	8%	49	5%	66	6%
Coming_up_with	Cognitive	2	17	4%	6	10	5%	49	5%	59	5%
Creating	Creative	2	11	3%	7	49	23%	172	18%	221	19%
Emotion_directed	Emotional	4	186	42%	18	16	8%	109	11%	125	11%
Experiencer_obj	Emotional	2	127	29%	26	28	13%	158	17%	186	16%
Intentionally_create	Creative	2	14	3%	9	55	26%	184	19%	239	21%
Subjective_influence	Cognitive	7	12	2%	6	8	4%	74	8%	82	7%

4.3 Degree and Value of Inspired Tweets to the Museum

Another output of discussing ‘inspiration-related’ FrameNet Frames with museum staff, in particular the *Subjective_influence* Frame, was the insight this provided into our research question regarding how the concept of inspiration might help museum staff think about relationships with visitors. At the point where candidate Frames were reviewed with museum staff, a potential model based upon the degree of inspiration emerged (Figure 3). This model was based on the ‘output’ FE slots from *Subjective_influence*, namely *action*, *behaviour* and / or *product*, any or all of which could be filled by the consequences of an individual being inspired.

Consideration of these three FE slots sparked a discussion about what the museum might achieve by inspiring its visitors. Firstly, two classes of *behaviour* were suggested: ‘everyday behaviour’ and ‘creative behaviour’. The example of everyday

behaviour discussed (shown in Figure 2) related to waste-disposal, a topic chosen because the museum was redesigning its nature displays and was thus focused upon environmental matters. *Creative* behaviour, however, is a key element of the strategy of Derby Museums: they aim to encourage their visitors' creativity as there is evidence that doing so benefits well-being (Carey 2006), and hence they valued inspiration with creative output more highly. The *Subjective_influence* FE slot *product* helped clarify thoughts about this – clearly a *product* is the outcome of being creative. Finally, the distinction between an *act* and a change in *behaviour* became an important part of the model. Therefore, the purpose of inspiring visitors, according to the model, is:

- (1) At the lowest level, to cause visitors to change the ways they performed everyday activities.
- (2) As 'inspiration increases', to cause visitors to try a new 'everyday' activity for the first time.
- (3) At the next level, to change visitors' everyday behaviour for good.

Then the museum may inspire visitors to:

- (4) Support pre-existing creative behaviour by encouraging them to try a different creative method.
- (5) As inspiration increases, to try a brand new creative act (e.g. learn a new skill such as throwing pots).
- (6) As inspiration further increases, to ingrain a new method into their creative technique in some way.
- (7) At the highest level, to cause them to embark upon a major life-change, such as a move to a creative career.

The potential evidence of inspiration found in Twitter data collected for this research predominantly belonged to types 4 and 5: people trying out new creative techniques or brand-new activities. Evidence of longer-term behavioural change, however, is likely to be harder to amass purely by finding potential expressions of inspiration in social media data, though potentially due more to ethical reasons regarding collection of personal data than for technical reasons (it is technically possible to download a lot of information about individuals over a long period from Twitter). A less invasive, more ethical method of establishing whether the life of a visitor had changed deeply would be to build up a long-term relationship with that visitor: indeed this type of relationship with the visitor would most likely need to happen in order to effect such a change anyway. One of the hopes museums have regarding social media is that it will help them to build and maintain such long-term relationships with their visitors (Russo 2011), though at time of writing there is currently little in the statistics and metrics systems provided by social media companies such as Twitter and Facebook to help monitor the building of such relationships.

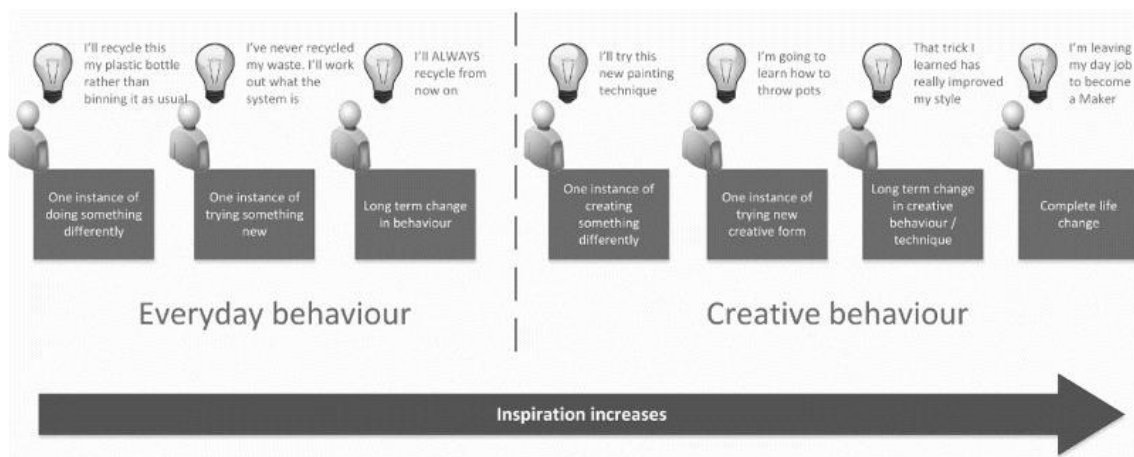


Figure 2 Emergent model based upon degrees of inspiration affecting change in visitors' behaviour

4.4 Exploratory Annotation of Related Tweets

The 1165 Tweets related to Frames via LUs were analysed by the first author to see which of the Frame Elements for the related Frame they contained, whether the correct sense of the related LU was used in the Tweet, and whether they contained a potential expression of inspiration. Some redundancy occurred as some of the Tweets were related to more than one Frame, because some of the LUs were repeated across different frames. For instance, the LU 'find' (and variations such as 'find out' and 'find myself') are part of the Frames *Becoming_aware*, *Coming_to_believe* and *Coming_up_with*. Under such circumstances, Tweets were annotated multiple times.

Certain features of the relationship between Tweets and their related Frames are worth mentioning. Firstly, only a small proportion of Tweets (11%) contained the terms required to fill all the Frame Element slots of their related Frame. This is unsurprising given the length of Tweets (i.e. max 140 characters), but in many cases, FE slot fillers could be inferred from the Tweet's context. For example, the following Tweet, related to the *Experiencer_obj* Frame via the *amaze* LU, is one of many from the dataset that was written in the first person:

So impressed with #museumsasconversations, absolutely amazing prototype.
#MMUK14 <http://t.co/wuz0fp2X5p>

In first-person Tweets such as this the sentient slot (the *Experiencer* in this case) is evidently the Tweeter. For second person Tweets (less common but still evident), the sentient is the reader, as shown by this Tweet, linked to the *Becoming_aware* Frame via the *spot* LU:

The lighting beams @derbysilkmill are drenched in history. When you are here try
& spot the makers mark #mmuk14 <http://t.co/prFoWNuOXm>

Both of these examples also illustrate a further important issue - Tweets are not purely textual; as both Tweets above contained images. Often, with instances containing media links, the filler of an FE slot could be inferred by looking at the accompanying image or video, such as with the Tweet shown in Figure 2. The linked image enables the

content of the *created_entity* FE of *Intentionally_create* to be filled meaningfully (with the filler ‘a museum exhibit’, shown in the image), though it would clearly be challenging to automate the process of filling this FE slot with such image-based content (see section 5). That relevant information can be overlooked by focusing exclusively on text when using Twitter as an evidence source was another of the issues noted by Tufeci (2014).

Another issue addressed by this annotation task concerned instances of LU links where the wrong sense of the LU appeared in the Tweet. FrameNet includes Part of Speech (POS) information for each LU, and some LU/ tagged Tweet POS mismatches were checked manually. A ‘by eye’ inspection indicated that there was a strong likelihood that incorporating POS information into the linking process would tend to improve linking accuracy.



Figure 3 Image attached to the Tweet provides a meaningful filler for FE slot ‘created entity’ of *Subjective_influence* Frame

4.5 Annotation of Tweets by Museum Staff

The second task involved the museum staff annotating Tweets by answering either yes or no to the closed question: ‘does this Tweet indicate that someone might have been inspired?’ This task was affected immediately by the issue illustrated in the previous section – i.e. making a decision about the quality of evidence based purely on text,

when often there were images or videos that contained potential evidence, simply made no sense to the museum staff. In order to test how ignoring pictorial information might impact the overall accuracy of the text-based Information Retrieval system, the decision was taken to allow the staff to view any linked media in order to support their yes / no decision.

The two members of staff were asked to annotate the data separately and their answers were compared: they agreed that Tweets constituted potential evidence of inspiration just over two-thirds of the time. The annotation session was recorded, and reasons for disagreement between the staff members were discussed. Because the Derby Museums staff were donating their time for free, the whole annotation session had to be concluded within two hours for 400 tweets, meaning that approximately one third of the cases upon which the staff disagreed could be discussed. A potential positive bias on the part of the staff was noted: on many occasions, the staff felt able to fill in surrounding context for Tweets based upon *their* experiences of the museum events, which had been a positive and inspirational experience *for them*. For example, the (typically sparse) Tweet:

After hours #senseofplates #MMUK14 <http://t.co/eX5hrZQHyO>

...was accompanied by a picture of a table containing some plates of cake and beer cans. Purely based on the content of the Tweet and image alone, there was no evidence of inspiration. However, to one of the museum staff, the Tweet illustrated the end of a ‘busy, happy day’ spent creating a prototype museum exhibit (the table the beer and cake are resting on was actually part of the prototype), and hence they thought it a valid expression of inspiration. There is, however, an argument against this particular Tweet as a ‘good’ expression of inspiration, given its potential to confuse people who did not attend the event, who would be unlikely to contextualize it in the same way.

To further investigate the potential impact of bias upon the annotation process, the first author annotated the same set of 400 Tweets annotated by museum staff, and, despite having attended both events, and having worked closely with Derby Museums over the previous six month period, the level of consensus between his opinions regarding which Tweets provided evidence of inspiration was 8% lower on average (between his opinions and that of each staff member) than that between the two members of staff (62.5% to 70.5%), and the researcher was much less generous in labelling Tweets as containing valid expressions of inspiration.

Once annotation was complete, precision and recall measures for the effectiveness of linking Tweets to Frame LUs as a method of retrieving likely expressions of inspiration could be calculated. A precision measure of 0.67 was recorded, alongside a recall of 0.35, resulting in an overall F-measure of 0.46. This is an encouragingly high precision and may imply that a relatively simple NLP system based on FrameNet performs well. However, the low recall indicates it could also have occurred because (as discussed) the staff were very generous in labelling Tweets as ‘containing expressions of inspiration’. This therefore resulted in a large number of false negatives (i.e. Tweets that the staff annotated as ‘good expressions of inspiration’ that did not contain an LU linked to a relevant frame).

4.6 *Generating a coding reliability confidence measure*

Table 3 shows the results of the first author’s reassessment of a sub-sample of the originally-annotated Tweets, 14 months after the original annotation. Column 2 shows the number of Tweets in the sub-sample, while column 3 shows the number of times the first author’s reassessment agreed with the original assessment. Columns 4 and 5 show the number of positive (i.e. those that contained a potential expression of inspiration) and negative annotations across both the original assessment and reassessment, while columns 6-9 show the nature of agreement or disagreement between the two assessments (e.g. the 0-0 column shows the occasions where the first author considered that the Tweet did not contain a potential expression of inspiration both times). Note the low number of 1-0 results (i.e. occasions where the researcher noted an expression of inspiration in the first annotation, but not in the second) compared with 0-1 results (when a Tweet was annotated positively the second time but not the first) – namely 9 to 44. This indicates that the researcher was a lot more positive (i.e. saw more potential expressions of inspiration) the second time.

The totals of disagreements were then combined with the overall number of positive results to generate Krippendorff’s Alpha for the comparison between both sets of annotations, resulting in an alpha measure of 0.52, which, while considerably better than chance, is far below the .800 recommended as a rule of thumb for complete coding reliability, and even considerably below the 0.67 result recommended for tentatively reporting preliminary results where it is acknowledged further work is required (Krippendorff, 2004:227). This low result was actually hypothesised given the lack of consensus between museum staff and the researcher observed while annotating Tweets collaboratively. The implications of this low measure of confidence in the reliability of the two annotations are discussed below.

Table 3: levels of intra-coder agreement by Frame

Frame	Sub sample size	# Agreements	Total +ve (1)	Total -ve (0)	0-0	0-1	1-0	1-1
Subjective_influence	20	17	31	9	3	3	0	14
Coming_up_with	13	8	9	17	6	5	0	2
Cogitation	19	16	15	23	10	3	0	6
Coming_to_believe	12	9	7	17	7	2	1	2
Becoming_aware	18	13	13	23	9	3	2	4
Creating	46	32	40	52	19	10	4	13
Intentionally_create	42	30	40	44	16	11	1	14
Emotion_directed	25	23	12	38	18	1	1	5
Experiencer_obj	32	26	26	38	16	6	0	10
	227	174	193	261	104	44	9	70

5 Discussion and Limitations

The research presented in this paper has shown that taking a novel automated computational approach using a rich semantic resource, such as FrameNet, in combination with a graph database, *has the potential* to perform well when retrieving potential expressions of inspiration. A high precision of 0.67 highlights that Tweet-based textual content can be quite reliably interpreted as containing expressions of inspiration, *where this was judged to be the case by the museum staff annotated dataset*. Future exploration of approaches to word sense disambiguation (in addition to using POS) would very likely increase precision significantly. It would also be a productive exercise to investigate and compare the effectiveness of different clustering algorithms (such as K-Means) or classification approaches (e.g. Random Forests). Nevertheless, the low recall of 0.35 highlights the problem that a lot of expression of inspiration was indeed missed by this approach, if indeed those being missed were expressions that did have genuine potential to denote inspiration, and were not just indications of overly-positive thinking by the museum staff conducting the annotation. Another key issue is that Tweets are not purely textual entities, so textual analysis of Twitter data is unlikely to provide comprehensive assessment of the evidence for a phenomenon.

A major caveat here, however, is the apparent susceptibility to bias of evidence based upon Twitter data, due to the contextualization (indeed the ‘framing’) of Tweets by those assessing their meaning. The sparsity of information in many Tweets tends to enable broader interpretation (indicated both by the intra-coder reliability measurement exercise, and by the first author’s direct comparison of his annotations with those of the staff), and this undoubtedly increased the impact of bias when annotation related to a phenomenon as complex as ‘inspiration’. Given that the initial definition of inspiration was based upon the working parts ‘emotion’, ‘cognition’ and ‘creativity’, perhaps the question: ‘does the Tweet indicate that the author was feeling, thinking or making anything?’ would have been less ambiguous and hence reduced the impact of positive bias.

Furthermore, the museums sector (indeed the broader creative arts sector) has approaches to dealing with bias in evaluation that could feasibly be applied to annotation exercises such as the one described here. When discussing the issue with the museum staff, they mentioned one such approach: Visit England’s Visitor Attraction Quality Assurance Scheme (VAQAS), which incorporates peer review into the evaluation process (Visit England 2015). ‘Annotation of training and testing data sets by peer’ might be another approach for making any automated system less positively biased.

The potential difficulty of designing an algorithm to fill slots in Frames from Tweet text automatically, and hence fully automate the process of finding potential expressions of inspiration, might be best illustrated by the following Tweet (linked to the *Subjective_influence* Frame via the *inspire* LU):

Brilliantly simple & effective #Engineering learning by @JWSYE at #DMMF14
@MakerFaireDERBY #STEM could take to @DerbyUK schools to inspire

Both museum researchers and the first author strongly agreed that the Tweet above was likely to be an expression of inspiration. Examining the context further, it

appeared the Tweeter was an educator, thus the Tweet seems to discuss taking ideas directly from the museum event to schools; further strengthening evidence of a potentially positive impact of the museum event. However, the syntax of this Tweet is idiosyncratic to the point that there was some debate about its exact meaning between the people analysing it, and it was not clear how the FE slots for the related *Subjective_influence* Frame might be filled using its text.

6 Conclusions

Fundamentally, this research and the research it built directly upon (Gerrard et al, 2014) indicates that the museum staff we consulted were still in the process of formulating a useful definition of ‘inspiration’. It is clear from our work that inspiration is a complex phenomenon with many ‘moving parts’, and hence it may be too complex a term to form a useful part of the core definition of museums such as that provided by the UK Museums Association. This is in part because the research also indicates that it is too complex a concept to structure Information Retrieval around such that the results are unambiguous and sufficiently free from bias to base trustworthy statistics upon. However, some of the thinking this research stimulated among museum staff upon the topic was sophisticated and potentially very useful: this was epitomised by the model proposed by the Derby Museums’ staff regarding ‘long-term’ inspiration and the potential for relationships with visitors to change their lives positively. While social media appears to have a role to play in the building of such relationships, it is clear that the social media platform used in our research (Twitter) did not provide the functionality required to monitor, support, evaluate and manage the building of such long-term relationships. Twitter is a platform based explicitly upon the immediacy of the here and now, not the future or the past, so while data from Twitter could feed into the type of Information System envisaged by our collaborators from Derby Museums, Twitter’s core platform does not provide much of the functionality to monitor longer-term, slow-building engagement that such a system would need. While it would be technically possible to evaluate how such engagement built outside of Twitter, using Twitter data, a large amount of visitor’s potentially sensitive and personal data would need to be collected, regularly, from Twitter, and stored elsewhere (with all the risk that entailed), for valuable, usable evidence of such relationship building efforts to emerge.

The first conclusion of this research, therefore, is that the datasets of Tweets related to the two museum events did contain potential expressions of inspiration that the staff from Derby Museums considered valuable evidence of the impact of their events upon visitors, and that between these datasets there was a substantial proportion of relevant Frames from FrameNet (24% Maker Faire, 22% MuseoMix). This indicates a positive answer to our first two research questions: inspiration is, potentially, a useful concept to help museum staff work with social media data productively, with some caveats discussed below.

However, one potential risk is that such *expressions of potential* inspiration do not constitute evidence that inspiration has *actually* occurred. There is a gap between expressions in written language and other online behaviours compared to the true experienced personal states of the individuals producing the social media content, and there is a danger that museum staff might equate such ‘expressions’ with the idea that ‘visitors had definitely been inspired’. Closing this gap is currently an area of active research (e.g. Sykora et al. 2015) and a real challenge for the broader field of opinion

mining (O'Connor et al., 2010; Ott, 2013; Eichstaedt et al., 2015). While the gap between real and expressed experience remains subject to further research, the best that systems such as the prototype developed for this research, or other sentiment analysis systems (many of which are commercially available), can offer are *indications of the possibility* that inspiration might have occurred.

The F-measure of 0.46 achieved from this study using 'out of the box' FrameNet Frames and Lexical Units indicates that there may be some potential to use FrameNet's Knowledge Patterns in social media data analysis. If more of the FrameNet data (such as the Parts of Speech of the Lexical Units) were used, and if efforts were made to annotate more informal text from social media using the same FrameNet Knowledge Patterns, results might improve considerably, particularly if the learnings taken from the critical evaluation of the annotation task described above were carried forward (i.e. base the annotation on more straightforward feeling / thinking / making questions, and incorporate peer review into the process). This indicates that we can answer our third research question regarding the value of the concept of 'inspiration' as potential Information Retrieval hook with a cautious 'yes', pending a substantial amount of further work using the 'deconstructed' parts of the concept.

The final issue regarding the effect of context and bias is the nature of Twitter data itself, which tends towards a sparsity that encourages broad interpretation: perhaps models of phenomena of the complexity of inspiration might be better suited to longer-format social media, despite the increased effort required to collect data from such sources. Conversely, IR systems based upon simpler concepts with fewer 'moving parts' than 'inspiration' might be better suited to retrieving information from Twitter: though it would be wise not to trust statistics based upon a Twitter IR system without conducting a coding reliability measurement exercise to check how consistently annotated any test reference sets used to train and evaluate the system in question were. However, it should also be noted that the nature of the museum events themselves, and the ways in which Twitter usage was encouraged at these events (MuseoMix in particular), will also have caused a positive bias in the data: i.e. Twitter users seem to have been more likely to Tweet positively about these events. This is not to say that the Twitter data was unrepresentative of their reactions, just that it tended to capture a particular class of reaction. The danger would lie in a failure to recognise this when analysing the data in question.

Also, Twitter's ephemeral nature does not lend itself to the monitoring of deeper relationships that build over time in ways that the Derby Museums staff suggested would be more supportive of the nature of inspiration in museums. At the very least, evidence of a phenomenon such as inspiration mined automatically from Twitter should be approached carefully: automating a system to collect such evidence in near-real-time is likely to be a risky approach that could potentially misinform. So while we can answer 'yes' to our first research question regarding whether the concept of inspiration helps museum staff think productively about relationships with visitors, we cannot claim that social media data (from Twitter at least) feeds neatly into this process at present. A semi-automated approach in which digital collection and analysis methods are used in a decision support role, and in tandem with other forms of qualitative analysis such as interviews with visitors, is a more prudent option currently.

Acknowledgements

Many thanks to Emma Hallam (Social Media Coordinator) and Jonathan Wallis (Head of Museums) at Derby Museums, and to Derby Museums as a whole. This work would not have been possible without the willingness to experiment which is fundamental to their organization. This research was partly funded by the UK Arts and Humanities Research Council: award number 1234317.

References

- Baeza-Yates, R. and Ribeiro-Neto, B., 1999. *Modern information retrieval*. Harlow, UK: Addison-Wesley.
- Baker, C.F. 2014. FrameNet: A Knowledge Base for Natural Language Processing. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929–2014)*, 1-5. New York, NY: ACL.
- Barrett, J. 2011. *Museums and the Public Sphere*. Malden, MA: Wiley-Blackwell.
- Baskerville, R.L. 1999. Investigating Information Systems with Action Research. *Communications of the AIS*, 2(3es), 4.
- Bertola F. and Patti V. 2016. Ontology-based Affective Models to Organize Artworks in the Social Semantic Web, *Information Processing & Management*, 52(1): 139-162.
- Brodie, R.J., Hollebeek, L.D., Juric, B. and Ilic, A. 2011. Customer engagement: conceptual domain, fundamental propositions, and implications for research. *Journal of Service Research*, 1,20 Thousand Oaks, CA: Sage.
- Cameron, F. 2008. Object-oriented democracies: conceptualising museum collections in networks. In *Museum Management and Curatorship*, 23(3), 229-243. New York, NY: Routledge.
- Carey, J. 2006. *What good are the arts?* Oxford, UK: Oxford University Press.
- Chang, S.K., Greco, L. and De Santo, A. 2015. Sentiment Detection in Social Networks using Semantic Analysis: a Tool for Sentiment Analysis and its Application in Cultural Heritage Realm, *Data Management in Pervasive Systems*: 183-193.
- Choudhury M. and Counts S. 2012. *The Nature of Emotional Expression in Social Media: Measurement, Inference and Utility*, Technical Report: Microsoft
- Clark, P., Thompson, J. and Porter, B. 2004. Knowledge patterns. In: Staab, S and Studer, R., eds, *Handbook on Ontologies*, 191-207. Berlin, Germany: Springer.

- Connolly, W.E. 2002. *Neuropolitics : Thinking, Culture, Speed*. Minneapolis, MN: University of Minnesota Press.
- Derby Museums. 2015. *Derby museums annual review 2014 - 2015*. Derby, UK: Derby Museums, 1. Accessed 14 June 2016. http://www.derbymuseums.org/wp-content/uploads/0135_Derby-Museums-Annual-Review-2014-15_24pp_lo-res.pdf
- Eichstaedt, J.C., Schwartz, H.A., Kern, M.L., Park, G., Labarthe, D.R., Merchant, R.M., Seligman, M.E. 2015. Psychological Language on Twitter Predicts County Level Heart Disease Mortality. *Psychological Science*, 26(2): 159–169.
- Fillmore, C.J. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences*, 280(1) 20-32. New York, NY: Annals of the New York Academy of Sciences.
- Finnis, J., Chan, S., & Clements, R. 2011. Let's Get Real - how to evaluate online success. Accessed 15 November 2013. <http://weareculture24.org.uk/projects/action-research/>: Culture 24.
- Fletcher A. and Lee M.J. 2012. Current Social Media uses and Evaluations in American Museums, *Museum Management and Curatorship*, 27(5): 505-521.
- Gerrard, D., Jackson, T., O'Brien, A. 2014. The Epiphany Project: Discovering the Intrinsic Value of Museums by Analysing Social Media. *MW2014: Museums and the Web 2014*
- Giaccardi, E. 2011. Things we value. *Interactions*, 18(1): 17-21.
- Goffman, E. 1974. *Frame Analysis*. UK: Harper and Row.
- Havasi, C., Speer, R. and Alonso, J. 2007. ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge. In *Recent Advances in Natural Language Processing*, 27-29. Philadelphia, PA: John Benjamins.
- Holden, J. 2006. *Cultural value and the crisis of legitimacy: why culture needs a democratic mandate*. London: Demos.
- Klein, H.K. and Myers, M.D. 1999. A set of principles for conducting and evaluating interpretive field studies in Information Systems. In *MIS Quarterly*, 23(1), 67-93. Minneapolis MN: University of Minnesota.
- Krippendorff, K. 2004. *Content analysis: an introduction to its methodology*. 2nd. ed. Thousand Oaks, CA; London: Sage.
- Langa, L.A. 2014. Does Twitter help museums engage with visitors? *ICConference 2014 Proceedings*: 484-495.

- Neo4J. 2013. Neo4j - the world's leading Graph Database. Accessed 30 December 2013. <http://www.neo4j.org>.
- Nuzzolese, A.G., Gangemi, A. and Presutti, V. 2011. Gathering lexical linked data and Knowledge Patterns from FrameNet. In Proceedings of the Sixth International Conference on Knowledge Capture, 41-48. New York, NY: ACM.
- O'Connor B., Balasubramanyan R., Routledge B., and Smith N. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series, Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, Washington D.C., USA.
- Ott, M., Cardie, C. and Hancock, J.T. 2013. Negative Deceptive Opinion Spam, HLT-NAACL Conference, Atlanta, USA.
- Pang, B. and Lee, L. 2008. Opinion Mining and Sentiment Analysis, Foundations and Trends in Information retrieval, 2(1-2): 1-135
- Ravi, K. and Ravi, V. 2015. A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications, Knowledge-Based Systems, 89(1): 14-46.
- Rippleffect. 2015. *Derby museums social media strategy*. Derby, UK: Derby Museums, 1. Accessed 30 July 2015. <http://www.derbymuseums.org/wp-content/uploads/Derby-Museums-Social-Media-Strategy-single-pages1.pdf>
- Russell, J.A. 1980. A Circumplex Model of Affect, Journal of Personality and Social Psychology, 39(1): 1161-1178
- Russo, A. 2011. Transformations in cultural communication: Social Media, cultural exchange, and creative connections. In Curator: The Museum Journal, 54(3), 327-346. Oxford, UK: Blackwell.
- Selwood, S. 2010. Making a Difference: the cultural impact of museums; an essay for NMDC. London: National Museum Directors' Council.
- Soren, B.J. 2009. Museum experiences that change visitors. In Museum Management and Curatorship, 24(3), 233-251. New York, NY: Routledge.
- Sykora M., Jackson T. W., O'Brien A. and Elayan S. 2013. Emotive Ontology: Extracting Fine-Grained Emotions from Terse, Informal Messages, International Journal on Computer Science and Information Systems, 8(2): 106-118.
- Sykora, M. D., Jackson, T., O'Brien, A., Elayan, S., & Von Lunen, A. 2014. Twitter based analysis of public, fine-grained emotional reactions to significant events. In ECSM - European Conference on Social Media, Porto, Portugal: ACPI.

- Sykora, M., Robertson, C., Shankardass, K., Feick, R., Shaughnessy, K., Coates, B., Haydn, L. and Jackson, T. 2015. Stresscapes: Validating Linkages between Place and Stress Expression on Social Media, ICML 2015 - 2nd International Workshop on Mining Urban Data, Lille, France
- Tay, D. 2014. Lakoff and the Theory of Conceptual Metaphor. In: Littlemore, J. and Taylor, J. R., eds, *The Bloomsbury Companion to Cognitive Linguistics*, 49-59. London, UK: Bloomsbury.
- Thelwall, M., Buckley, K. and Paltoglou, G. 2012. Sentiment Strength Detection for the Social Web, *Journal of the American Society for Information Science and Technology*, 63(1): 163-173.
- Tufekci, Z. 2014. Big questions for social media big data: representativeness, validity and other methodological pitfalls. In *ICWSM '14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, Paulo Alto, CA: AAAI.
- UK Museums Association. 2014. Museums Association FAQs, Available: <http://www.museumsassociation.org/about/frequently-asked-questions>. Accessed 6th January 2014.
- Villaespesa, E. 2015. An evaluation framework for success: capture and measure your social-media strategy using the balanced scorecard. *MW2015: Museums and the Web 2015*
- Villaespesa, E. 2013. Diving into the museum's Social Media stream. Analysis of the visitor experience in 140 characters. In *Museums and the Web 2013: Proceedings*. Archives & Museum Informatics.
- Visit England. 2015. Quality assessment for your attraction. Accessed 27 July 2015. <https://www.visitengland.com/biz/advice-and-support/businesses/quality-assessment-your-attraction>