Article

Analyzing User Digital Emotions from a Holy versus Non-Pilgrimage City in Saudi Arabia on Twitter Platform

Kashish Ara Shakil [1], Kahkashan Tabassum [1], Fawziah S. Alqahtani [1] and Mudasir Ahmad Wani [2,*]

1   Department of Computer Science, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh Saudi 11671, Saudi Arabia; kashakil@pnu.edu.sa (K.A.S.); ktOsman@pnu.edu.sa (K.T.); fsalqahtani@pnu.edu.sa (F.S.A.)
2   Department of Information Security and Communication Technology, Norwegian University of Science and Technology (NTNU), 2815 Gjøvik, Norway
*   Correspondence: mudasir.a.wani@ntnu.no

Abstract: Humans are the product of what society and their environment conditions them into being. People living in metropolitan cities have a very fast-paced life and are constantly exposed to different situations. A social media platform enables individuals to express their emotions and sentiments and thus acts as a reservoir for the digital emotion footprints of its users. This study proposes that the user data available on Twitter has the potential to showcase the contrasting emotions of people residing in a pilgrimage city versus those residing in other, non-pilgrimage areas. We collected the Arabic geolocated tweets of users living in Mecca (holy city) and Riyadh (non-pilgrimage city). The user emotions were classified on the basis of Plutchik's eight basic emotion categories, *Fear*, *Anger*, *Sadness*, *Joy*, *Surprise*, *Disgust*, *Trust*, and *Anticipation*. A new bilingual dictionary, AEELex (Arabic English Emotion Lexicon), was designed to determine emotions derived from user tweets. AEELex has been validated on commonly known and popular lexicons. An emotion analysis revealed that people living in Mecca had more positivity than those residing in Riyadh. Anticipation was the emotion that was dominant or most expressed in both places. However, a larger proportion of users living in Mecca fell under this category. The proposed analysis was an initial attempt toward studying the emotional and behavioral differences between users living in different cities of Saudi Arabia. This study has several other important applications. First, the emotion-based study could contribute to the development of a machine learning-based model for predicting depression in netizens. Second, behavioral appearances mined from the text could benefit efforts to identify the regional location of a particular user.

Keywords: emotion analysis; Twitter; online social network analysis; behavioral analysis

1. Introduction

Emotions are a vital source of information to assess the overall psychological well-being and health of individuals. Emotion plays an important role in the overall behavior and reactions of an individual in different circumstances. Findings have shown that an individual's response to natural surroundings has an impact on their parasympathetic nervous system. The surroundings also have a positive impact on a person's emotional state and psychological activity level [1]. Thus, people living in varying regions will have a different emotional state. The traditional methods for studying these individual states involved questionnaires, interviews, or in-person observations. However, such studies to date generally have been marked by small sample sizes and involved labor-intensive efforts in reaching out to individuals, leading to inconsistency and inaccuracy in the inferences obtained. In the present-day world, social media sites such as Twitter, LinkedIn, and Facebook are increasingly becoming platforms for individuals to express their emotions, opinions, and ideas. These days people are more expressive on such platforms, as they can fully express themselves without revealing their real identities or endure the stigma of

being judged. Thus, digital footprints allow social media platforms to serve as a rich source of information to investigate the emotional dimensions, sentiments, and psychological behavior of individuals. These data, accumulated over time, have led to granular emotion studies for every user. For example in one of the studies, tweets were used to analyze anti-LGBT campaigns in Indonesia [2]. In another work, Twitter messages were used to analyze dialogue during the Austrian presidential elections in 2016, concluding that the winning candidate sent tweets that were neutral in comparison to his opponent, whose messages were based on emotions [3].

Saudi Arabia is a country that is transforming itself into a social media powerhouse. Of its total population of 34.54 million, 25 million are active social media users, and of these, 20.03 million are Twitter users [4]. With a majority of the country's population being avid Twitter users, we can use the data they share to understand how user environments affect their overall mental wellbeing and behavior. To gauge the emotional atmospheres of residents, we studied the tweets shared by users in Saudi Arabia hailing from different backgrounds. In particular, we studied the tweets of users from a holy city or place of pilgrimage i.e., Mecca, and a non-pilgrimage metropolitan city i.e., Riyadh. This work sought to show the potential of social media data shared by users in uncovering the impact of users' surroundings on their mental health. Another goal was to show how their emotions vary across different social environments.

At present most of the related work on identifying emotions on social media has focused on assessing user sentiments. Approaches such as supervised deep learning, semi-supervised methods, natural language processing, and lexicon-based analysis have been used to detect user sentiment polarity of social media data [5–7]. Due to the increased interest of users in performing analysis of social media conversations, a new field of emotion analysis has emerged [8]. In the current literature, studies have been carried out to understand and detect different forms of emotions such as sarcasm [9], rumors [10], depression [11–13], and suicidal tendencies [14]. However, there are primary limitations on literature research in this field. Given the thrust of analyzing user emotions, there is a need to focus on understanding the basic emotion categories, which act as a foundation toward effectively understating emotional behaviors. Deriving the emotions of users rather than their sentiment polarities (positive and negative) will enable us to investigate the digital emotion footprint of users. Therefore, to gain deeper insight into specific emotion dimensions and categories, we focused on the study of basic emotion categories proposed by [15]: namely, fear, trust, surprise anger, sadness, joy, disgust, and anticipation. We designed a bilingual (English and Arabic) emotion lexicon named AEELex based on the Moodbook [16] and NRC Word-Emotion Lexicon (Emolex) [17]. We have extended these dictionaries for the Arabic language, along with new words from user tweets, to form our bilingual dictionary. Following are the main contributions of our work:

- The geolocated tweets from two cities in Saudi Arabia have been extracted using Twitter API (the dataset will be made available with the algorithms for the researchers working in this domain).
- AEELex, a bilingual emotion lexicon, has been designed to extract the different emotions of users in two different (pilgrimage and non- pilgrimage) cities.
- The potential of user tweets to harness user emotions has been studied.
- The different emotion-expressing behaviors of netizens across two different cities has been analyzed and compared.
- The most common emotion-based vocabulary terms (both Arabic as well as English) belonging to eight emotion categories from both user groups have been identified.

The rest of this paper is organized as follows. Section 2 presents related work. Section 3 describes the approach we adopted for our study. Sections 4–6 describe the data collection, lexicon design, and lexicon comparison and validation respectively. Section 7 presents a data analysis. Section 8 is a discussion on the inferences from the study. Section 9 contains the conclusions of the paper, outlining potential research directions and the contributions of this study.

2. Related Work

Social media data have been used in many applications across different domains for various purposes, including detecting humor in social media slang used on platforms such as Weibo, the largest Chinese social network [18], or analyzing the spatial patterns of emotions expressed by users visiting theme parks such as Disney World within the dimensions of high and low arousal or pleasure and displeasure [19]. Several studies have been performed to identify user satisfaction based on positive emotions such as happiness and pleasure [20]. Thus, user emotions vary with changes in their region and are also reflected in their social media accounts. One such study showed how the emotions and sentiments of users toward climate change vary between different countries such as the U.K. and Spain. It showed that Twitter users from the U.K. talked less negatively about the weather in comparison to Spanish citizens and had anticipation as their predominant emotion [21]. In [16] users' emotions were analyzed in a region in conflict, Kashmir, and a region marked by the absence of conflict, Delhi, two union territories in India. The authors showed that people residing in to conflict-impacted regions have a negative attitude and express negative emotions such as anger, fear, and sadness in their Facebook posts. The proposed study also analyzed users' emotions across different regions in two different cities in Saudi Arabia. Furthermore, to bridge the gap between different nationals and tourists residing in a given region, and for more accurate analytics, we used data that were in the native language of the country i.e., Arabic.

Most of the work performed with tweets from Saudi Arabia has focused on sentiment and semantic analysis [22,23]. Words were labeled based on their polarity. In [24] SANA, a multi-dialect, multi-genre, and multilingual lexicon, was created based on text from the Egyptian chat room on Yahoo Maktoob. The words here were labeled based on their positive, negative, and neutral content. In [25], a study was performed on social media data collected from YouTube, Facebook, Al-Saha Al-Siyasia, and Al Arabiya based on the 2009 and 2011 floods in Jeddah. The study showed that people on YouTube posted more emotional comments than those on Facebook. The users expressed emotions such as sadness for the loss of lives and anger toward authorities. Saudi Mood [25] is a real-time tool for visualizing the emotions of users on Twitter, covering happy, sad, angry, scared, and surprised emotions only. The authors in [26] manually created a dataset of Arabic emotions using tweets, but the tweets were filtered from Egypt geolocation only. The emotions in this case were placed into the categories of happiness, joy, sadness, anger, disgust, fear, and surprise. Additionally, the data were categorized on the basis of only 5879 tweets. It has been observed that there are very few studies based on the Plutchik [15] emotion categories in the Arabic language. Most of the work in this discipline has been conducted using either sentiment or emotion analysis. However, the studies on emotion analysis have been limited either by their size or the emotion categories they assessed. This study is a maiden attempt to analyze the difference in emotions of users residing in a pilgrimage area versus those in a non-holy city. An analysis of this kind can assist in designing a machine learning-based system to predict the mental state of a user and the location of a tweet/post.

3. Approach Synopsis

Our analysis involved four important steps, as shown in Figure 1. In particular, we performed an analysis of the emotional behavior of people belonging to two different regions. The emotional behavior in this context could be defined as the choice of emotion words used by individuals to express their feelings and opinions.
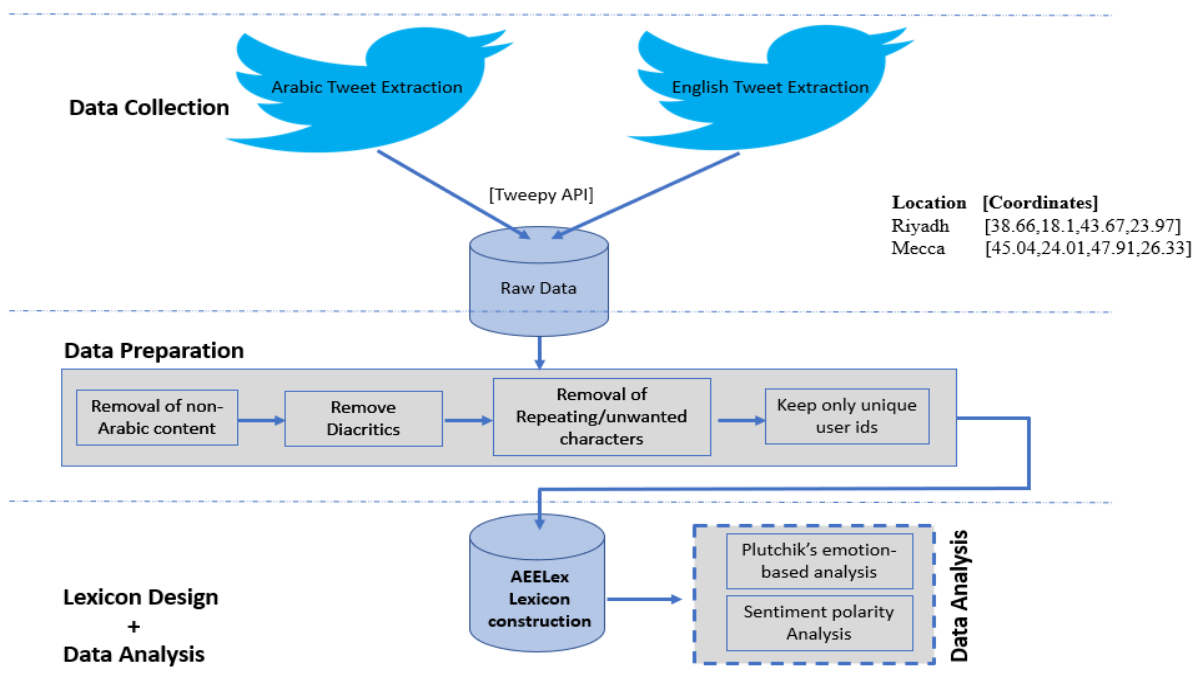
Figure 1. Approach synopsis for studying holy city versus non-holy city.

4. Data Collection

Data were collected through filtered real-time streaming API [27] provided by Twitter. User tweets were collected based on their locations and language, excluding retweets. The data were harvested from selected Twitter profiles for a span of 1.5 months, with an average of 20 tweets collected per user. The collected data did not take into account the activeness of the users. The tweet language was Arabic. In particular, we focused on Saudi dialects including Najdi Arabic, Gulf Arabic, etc. The users' data were collected from two metro cities of the Kingdom of Saudi Arabia, i.e., Riyadh and Mecca. The regions were selected using the bounding box option that allows geo-tagged user tweets to be collected based on a user-defined bounding box area. The geolocation co-ordinates of Riyadh (45.04, 24.01, 47.91, 26.33) and Mecca (38.66, 18.1, 43.67, 23.97) were obtained using Klokantech bounding box tool [28]. The choice of these cities helped us in studying the contrasting emotions of people belonging to a place of pilgrimage, and a city within the same country but not a place of pilgrimage. Table 1 shows the statistics of the data collected.

Table 1. Statistics of the data before preprocessing.

Item	Size
Total number of users	85,027
Total number of tweets (~20 tweets per user)	1,700,540
Tweets collected from Riyadh	951,600
Tweets collected from Mecca	748,940
Number of distinct users in Riyadh	47,580
Number of distinct users in Mecca	37,447

5. Arabic English Emotion Lexicon (AEELex) Design

AEELex is a bilingual emotion lexicon for extracting user emotions. This lexicon has been developed in Arabic and English language. This section discusses the steps for data preparation and design of the proposed lexicon.

5.1. Data Preparation

At this stage, the data were prepared to suit the needs of the experiments that would be carried out for performing analysis. We collected more than 85K tweets. Without being preprocessed, these tweets were highly redundant and contained a large amount of repetitive information, such as multiple tweets by the same users, etc. To address this issue, the geolocated tweets from Saudi Arabia were collected and cleaned using the following steps.

1. *Exclude redundant tweets*: During this step, all of the redundant tweets were removed from the data.
2. *Remove diacritics:* The script of the Arabic language has many diacritics. These diacritics or consonants help in modifying the pronunciation of the language [29].
3. Arabic diacritics are short vowel symbols that are optionally written on the top or under the Arabic letter to guide the intended pronunciation. There are many Arabic diacritics, but the basic and most used are Fathan (◌َ), Damma (◌ُ), Kasra (◌ِ), Tanwin Fath (◌ً), Tanwin Damm (◌ٌ), Tanwin kasr (◌ٍ), Sukun (◌ْ), Shadda (◌ّ), and Tatwil/Kashida (-), which are described briefly in Table 2 [30].
4. These Arabic diacritics are mostly used in classic Arabic scripts such as in the Quran and Hadith, their quotes, poetry, children's literature, or when a word has many different meanings according to its diacritics. However, adult native Arabic speakers can easily distinguish a word's meaning based upon the word context. Generally, these diacritics are considered noise that needs to be removed. Therefore, in this work, all the diacritics mentioned in Table 2 were removed.
5. *Remove repeating characters:* At this step, we checked our data, and if a character appeared more than once in a word, then the repeated character was removed. For example, if the word contained repeated characters such as "مبروووووووك", with the character "و" in this case repeated several times, we removed the repeated character and rewrote the word as "مبروك", meaning congratulations.
6. *Remove punctuation marks:* All punctuation marks such as '÷ × ¦ <> _ ( )*&^%] [—"..."!¦ + ¦ ~ {}' ,.؟":/‹ - were removed.
7. *Remove non-Arabic words*: Because the focus of this study was based on Arabic tweets, only Arabic words were kept.
8. *Keep only unique user IDs while preserving their tweets:* Within the data, many of the users had posted more than one tweet. Therefore, only unique users were retained but their respective tweets were merged and preserved for analysis. The schema for the final data that were used for the experiments is shown in Table 3.

Table 2. Information on the main Arabic diacritics (diacritic name, shape on the letter "م", and sound).

Name	Shape	Sound
Fatha	مَ	a
Damma	مُ	u
Kasra	مِ	i
Tanwin Fath	مً	an
Tanwin Damm	مٌ	in
Tanwin kasr	مٍ	un
Sukun	مْ	No sound
Shadda	مّ	Doubling or extra length
Tatwil/Kashida	-	No phonic value

Table 3. Schema Used in Experiments.

Unique User ID	Aggregated Unique Tweets	Region (Mecca/Riyadh)

Table 4 shows statistics on the data after data preprocessing. Furthermore, to make our data ready for analysis, Natural Language Toolkit (NLTK) was utilized to remove stop words (Arabic language) and tokenization. Additionally, neutral tweets were omitted, and only tweets showing positive or negative polarity were used.

Table 4. Statistics on the data after preprocessing.

Item	Quantity
Total number of tweets	35,383
Tweets collected from Riyadh	16,726
Tweets collected from Mecca	18,657

5.2. Lexicon Design

The lexicon-based analysis is one of the main approaches used in sentiment analysis [31]. It includes identifying the sentiments of users based on the semantic orientation of their words [32]. This method relies on a dictionary-based approach comprising a list of words according to their semantic orientation. There are different methods of creating such dictionaries, including automatic [33] and manual [34].

For the current study, a bilingual emotion dictionary comprising emotions expressed in the English and Arabic languages was created. This dictionary is named "AEELex" and is based on eight emotions proposed by Plutchik [15] including *Fear* (خوف), *Anger* (غضب), *Sadness* (حزن), *Joy* (فرح), *Surprise* (مفاجأة), *Disgust* (اشمئزاز), *Trust* (ثقة) and *Anticipation* (ترقب). AEELex comprises these eight emotion categories along with two sentiment polarities, i.e., positive (إجابي) and negative (سلبي).

Due to the complexity of the language used, a manual approach was adopted with the involvement of native Arabic language experts. The emotions fear (خوف), anger (غضب), disgust (اشمئزاز), and sadness (حزن) are associated with negative sentiments, whereas joy (فرح), trust (ثقة) and anticipation (ترقب) fall in the category of positive sentiments. The category surprise (مفاجأة) is a mixed sentiment comprising both positive and negative emotion words.

Apart from the words belonging to the eight mood categories, additional words were manually added under the category of positive and negative emotions. We developed our dictionary based on the Moodbook and NRC Word-Emotion Lexicon (Emolex). Both Emolex and Moodbook also tag words according to Plutchik's eight emotions along with the two categories of semantic polarities—positive and negative. However, neither dictionary categorized the moods according to our context. Moreover, these dictionaries are based on words in the English language only. Therefore, we extended these dictionaries to the Arabic language, along with new words, to form our bilingual dictionary, "AEELex". This dictionary was developed using more than 19,000 tweets written in the Arabic language. Table 5 shows an example of a few mood words in the AEELex dictionary.

Table 5. Example of emotion words in AEELex.

Mood Category (English)	Mood Category (Arabic)	Mood Words (English)	Mood Words (Arabic)	Example Sentence (From the Collected Dataset)
Fear	خوف	afraid	يخاف	يخاف تطلعه من البيت
		darkness	ظلام	لايوجد شخص في الدنيا يساوي مكانه
				الوالدين وأثر فقدهم على الشخص ومع ذلك تستمر الحياه ولا ظلام ولا يأس
		trembling	ارتجف	حتى وان ارتجف صوتي أسمعني للنهايه
Anger	غضب	wild/brutal	وحشي	رسل له الفيديوهات التي تثبت وحشيتهم ضد الاسرئ
		war	حرب	حيران مشبوه المضاجع ليله حرب وليل النائمين سلام
		cruel	قاسي	قاسي علي اليوم
Sadness	حزن	pain	الم	.أميل للبقاء حول الإنسان المتفهم
				الإنسان الذي يقدر ظروفك ، الم قلبك ، تكدر مزاجك وإرهاق روحك
		mistake	غلطه	وهالمره غلط غلطه ولكني عجزت اغفر. تعودت الرضا منه ولا قصر معي مره
		broke down	كسر	مجر الرياض الطبيعي مع كسر الرخام
Joy	فرح	fun	مرح	هيا بنا !وقت مرح
		dance	رقص	هه محرم وليد الفراج رقص العنزي على واحده ونص يقول له اسأل رئيس النصر
		entertainment	ترفيه	مب صحيح كلامك ، وزاره الرياضه ما راح تدخل بالملعب بيكون للنصر وهيئه الترفيه فقط
Surprise	مفاجأة	wondrous	عجيب	نسأل الله اللطف والحمايه يارب فعلا إسراف شي عجيب
		oh god	يالله	يالله برزق مالحد فيه منه ينسي المحتاج وش كان يحتاج
		a daemon	شيطان	توضاء وتعوذ بالله من الشيطان ودايم ادعي الله انه يوسع
Disgust	اشمئزاز	disgusting	مقزز	ما يحتاج اتثقف على شيء مقزز
		bullshit	هراء	شخص في الجيش لماذا هذا الهراء
		disadvantage	عيب	عيب لا تعممي افكارك على كل الرجال
Trust	ثقة	certain	متأكد	متأكد ؟ مش الهاملاين ترا
		support	دعم	على الاقل .تقديم دعم مالي من احمد كعكي .بمبلغ مليون يشكر عليه
		feeling	إحساس	في إنو أكثر الأشياء الي إحساس الإمتنان له
Anticipation	ترقب	prospect	احتمال	ولد عمي يدرس في شرق اسيا يقول فيه اشاعات ان فرق شرق اسيا
				بتنحسب بسبب انتشار الفايروس عندهم بشكل خطير و احتمال يتوجهوا
		anxiety	قلق	اللهم حيث لا قلق ولا هم ، اللهم حيث الراحه دائما
		feel	احس	احس استحيت

6. Lexicon Comparison and Validation

We compared our bilingual lexicon with two recently proposed emotion lexicons, one in the Arabic language proposed by M. Saad [35] and the second one, an English language lexicon, Moodbook. Table 6 highlights the characteristics of these two lexicons along with the proposed AEELex.

M. Saad's Arabic language lexicon comprises six categories of emotions: namely, *Fear*, *Anger*, *Sadness*, *Joy*, *Surprise*, and *Disgust*. However, AEELex has two additional categories i.e., trust and anticipation. Furthermore, AEELex has a comparatively larger number of emotion words than the former. Therefore, the proposed lexicon provides a much more fine-grained study of emotions and can cover many numbers of tweets where the existing lexicons failed to mine any emotions. We performed a comparative analysis by identifying the emotion categories of 1000 tweets in the Arabic language collected from two cities in Saudi Arabia and applying it to AEELex (Arabic version) and the lexicon proposed by M. Saad. It was observed from the results that both lexicons were able to identify all the emotion categories. However, M. Saad's lexicon was not able to mine emotions from ~60 percent of total tweets; contrastingly, AEELex was not able to mine only 17 percent of total tweets. M. Saad's lexicon combined several emotions into a single category, which led to a lack of essence for particular emotion words such as *Joy*, *Trust*, and *Anticipation*— emotion categories that were combined into one category (*Joy*). For example, the emotion

word " عبادة"(worship) was placed under the *joy* category by M. Saad, whereas AEELex classified the same word under the *anticipation* category rather than *joy*.

Table 6. Comparison of AEELex with other lexicons.

Lexicon	# Emotion Categories	Distinct Emotion Categories Identified	Positive/ Negative Polarity	# Tweets Where Lexicon was Not Able to Find Any Emotion
M. Saad lexicon [35]	6 (Fear, Anger, Sadness, Joy, Surprise, and Disgust)	6	No	595
AEELex (Arabic)	8 (M. Saad lexicon + Trust and Anticipation)	8	Yes	177
Moodbook	8 (Fear, Anger, Sadness, Joy, Surprise, and Disgust, Trust and Anticipation)	8	Yes	732
AEELex (English)	8 (Fear, Anger, Sadness, Joy, Surprise, and Disgust, Trust and Anticipation)	8	Yes	663

means 'number of'.

Furthermore, we compared AEELex with an English language lexicon, Moodbook, by analyzing 1000 tweets in the English language collected from Saudi Arabia. Both lexicons have eight emotion categories and were able to identify all the categories in the user tweets. However, it was observed that AEELex outperformed Moodbook by being able to identify a greater number of emotions in the tweets.

Thus, we can conclude that AEELex is efficient as a bilingual dictionary in identifying more emotions from text in both English and Arabic.

7. Data Analysis

We carried out a study to identify differences in the emotional state of Twitter users residing in Riyadh and Mecca. We aggregated all the tweets of users and analyzed them using AEELex. The analysis was performed in two phases. The first phase showed that tweets had the potential to harness their users' emotions. In the second phase, we highlighted differences in emotions shared by people residing in a religious city or holy city (Mecca) versus a metro city (Riyadh)in Saudia Arabia.

7.1. Phase 1: Sentiment Polarity Analysis (Positivity and Negativity)

The objective of this phase was to determine if the user tweets revealed the peace and positivity experienced by residents living in a holy place. Therefore, we compared the amount of positivity and negativity in user tweets. AEELex contains a list of positive and negative emotion words. *Joy*, *Trust*, and *Anticipation* contain words belonging to the positive category whereas *Fear*, *Sad*, *Angry*, and *Disgust* comprise words falling under the negative category. Words belonging to the *Surprise* category were classified based on the context. For each region, we calculated the positivity and negativity among the users with Equations (1) and (2), respectively.

$$pos_{region} = \frac{PositiveEmotionWords(region)}{TotalEmotionWords(region)} \tag{1}$$

$$neg_{region} = \frac{NegtiveEmotionWords(region)}{TotalEmotionWords(region)} \tag{2}$$

where, pos_{region} and neg_{region} are the positivity and negativity of a region, respectively. *PositiveEmotionWords(region)*, *NegativeEmotionWords(region)* are the number of positive

and negative emotion words of a region, respectively, under eight emotion categories and $TotalEmotionWords(region)$ is the total number of emotion words extracted from the tweets of a user belonging to that region.

Furthermore, the plots in Figure 2 show the proportion of positive and negative feelings expressed by people residing in the cities of Mecca and Riyadh. It is observed from the plots that users from both regions expressed positive emotions much more than negative ones. However, upon conducting the experiments (to support our comparative study), we observed that users from Mecca city expressed a high degree of positivity (~90%) in comparison to the positivity expressed by users residing in Riyadh (~70%). This can be attributed to the fact that people residing in holy places are regularly exposed to religious remembrance and rituals that inculcate a shared feeling of inner harmony due to spiritual enlightenment and inner satisfaction. Conversely, people residing in a metropolitan city such as Riyadh live a fast, competitive, stressful, and busy life. The positivity of Mecca residents could be seen from their tweets about God and friendship and positivity such as " ما شاء الله عليك بس الله يحميك " (As God wills, but God will protect you), " الله بيسر امرك " (May Allah make it easy for you), " بالتوفيق يا صديقي " (Good luck my friend), etc. The tweets from Riyadh usually talked about competitive affairs of day-to-day life such as " ایش آره مال منم " (What kind of money is this?) and " الجو اليوم عليل " (The weather today is fine).
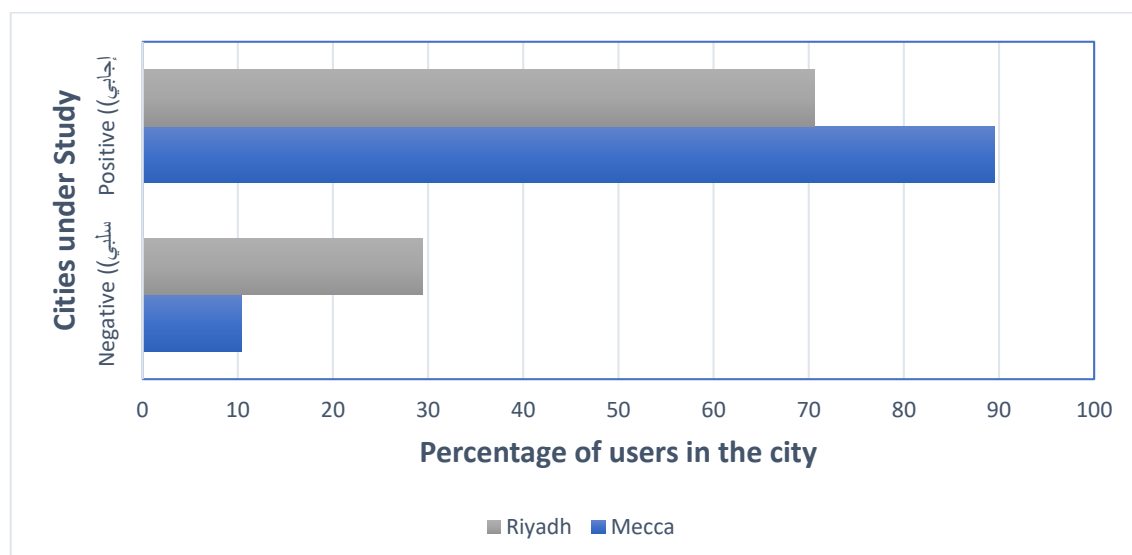

Figure 2. Proportion of positive and negative feelings expressed by people residing in Mecca and Riyadh.

Hence, from this study, we could perceive that user tweets have the potential to show the variable emotions of residents of different areas and consequently the emotional wellbeing of their users. This work thus proffers analysis of user tweets as a method of perceiving the mental state of a person.

7.2. Phase 2: Plutchik's Emotion-Based Analysis

This phase investigated how the emotions of users changed across the two regions. The varying emotions of the users were observed across different mood categories based on Plutchik's emotion wheel, as shown in Figure 3. Figure 4 illustrates how the proportion of user moods differed across the two regions. In the category of the most expressed emotion, which was *Anticipation*, the users from Mecca were ahead (50%) in comparison to those from Riyadh (around 40%). The users in Mecca tended to use words such as " الله " meaning "God suffices" in their tweets: for example, " استقبال الضيوف الشيخ عبدالله الغنيم يرحمه الله " (Reception of guest Sheikh Abdullah Al-Ghunaim, may God have mercy on him). As shown in the plot, users from both the regions Riyadh and Mecca experienced *Joy*, but a higher

proportion of the users in Mecca, approximately 20%, expressed the emotion *Joy*. In this city, the users use words such as " ورد" (flower), "حب" (love), etc. in their tweets: for example, " انا بس تخيلت انه ورد ازرق روقت" (I just imagined that it was blue roses) expressing positivity and happiness, " انا اقدر الصداقات العميقه جميله مليانه حب") I appreciate deep, beautiful friendships full of love), etc.

Figure 3. Eight emotion words [37].

Figure 4. The proportion of user moods across Riyadh and Mecca.

For emotions falling in the negative categories, *Sadness* (حزن) was the one most expressed. Users from both regions used words such as "تعبان" (tired), " الم" (pain), etc. to express sad emotions; for example, tweets such as " تعبان شكلها فره التيك توك قبل النوم" (I am tired, it seems that because I spent a long time on Tik Tok before bed), " لا تكن المضحي دائما ، تمرد ," كن الشخص السيء لمره واحده" (Don't always bear the pain, pass by, be the bad person for once) have been seen on user profiles. *Anger* (غضب) was the second most expressed emotion in this category, articulated with phrases such as "جريمه" (crime), " جوع" (hungry) in tweets such as " السرقة جريمة" (stealing is a crime), "جوع آخر الليل ماندري وش نسوي فيه" (Late night hunger, what do we do?), followed by *Disgust* (اشمئزاز) expressed with words such as "عيب" (shameful), " تافه" (petty), etc., in tweets such as "عيب عيب محتشمه لاترسموها زي كذا" (Shame, shameful, don't draw it like this), "تافهة كان يمكنني فعل اكثر ما تترقب" (Petty, I could have done more than you expect). It was also shown that the proportion of users falling

into these negative categories was much higher in Riyadh than Mecca. These persistently negative feelings inculcated in a population can lead to psychological issues such as bipolar disorders, depression, etc. It has been reported that psychological ailments such as drug use, depression, and anxiety-based disorders are among the leading causes of disability in Saudi Arabia [30]. Furthermore, according to a study [36], as drugs, petrol, lighter fluids, and tobacco were considered to be easily accessible to young generations in the city of Riyadh. This could be one of the reasons why the people of Riyadh show more negative and depressive emotions in their tweets in comparison to people in Mecca city.

8. Discussions

In this paper, we made use of social media platforms such as Twitter to study the emotions of people residing in a holy city versus those residing in a typical metropolitan center in the same country. Twitter was used as a choice of social media in this study, as this work was focused on the extraction of sentiments (emotions) from text. The Twitter network appeared to be one of the main sources of text through which users express their feeling in just 280 characters. Other social networking platforms such as Facebook hold other user content (audios, videos, images in almost equal ratios) as well. Although the inclusion of all the different types of media content was beyond the scope of this study, it could be used in our future work. To study differences in user emotions between the two cities, we performed various experiments and conducted a comparative study based on users in these cities. To perform this comparative study, we considered tweets posted by users in Mecca (pilgrimage region) and Riyadh (typical metropolitan city).

Our findings indicate that people residing in Mecca and Riyadh have significant differences in their emotional states. The entire study was divided into two phases. In the first phase, we concluded that tweets posted by users have the potential to reveal their expressed emotions and thus their mental states. People from holy areas expressed much more positivity than citizens from a typical city. The hectic and stressful life of the city affects people negatively, leading to comparatively more psychological disorders and negative feelings such as anger, disgust, and sadness expressed in words such as "عيب" (shameful), "الم" (pain), or "تافه"(petty).

Conversely, people residing in religious places such as Mecca, one of the holiest cities in the world for Muslims, are much more at ease and at peace with their life. They tend to turn to God (Allah) when in need and as a result shut out their negative emotions. Hence, they are much happier and have a positive outlook on life. They tend to use spiritual words such as "الله" (Allah), "الحمدلله" (alhamdulillah meaning thanks to Allah), and "يارب" (ya Rabb meaning O Lord) revealing their emotional connection with god.

In the second phase, we performed a fine-grained analysis of user emotions and studied the differences in emotions shared by people residing in the two cities. We observed that words in the positive categories such as Joy and Anticipation were used by a higher proportion of people in Mecca (~90%) compared to those in Riyadh (~70%). In order to perform a more fine-grained analysis, we created a word-cloud of the words extracted from the tweets of people residing in Mecca and Riyadh cities, as shown in Figures 5 and 6, respectively. For a better understanding of user emotions, the top five words in each of the emotion categories expressed by users in the two cities are presented in Table 7. It can be seen that some of the words used are common in both cities, such as "خوف" (meaning fear) and "خلاص" (salvation/finish) and rank at the topmost position in their corresponding categories, i.e., fear and anger, respectively.

Table 7. Top 5 words from each emotion category.

| Mood Category	Riyadh	Riyadh (English)	Mecca	Mecca (English)
Fear	خوف	fear	خوف	fear
	خايف	scared	جن	genie
	قلق	worry	خايف	scared
	يخاف	are afraid	يخاف	are afraid
	أخاف	I'm afraid	جور	yore
Anger	خلاص	Salvation/finish	خلاص	Salvation/finish
	صعب	difficult	مر	bitter (noun)
	مر	bitter (noun)	صعب	difficult
	غبي	stupid	غالي	expensive
	كره	hate	عيب	fault
Sadness	هم	worry	هم	worry
	شد	tightened	الم	pain
	الم	pain	شعور	feeling
	شعور	feeling	حزن	sad
	طلب	request	فقد	lost
Joy	الحمدلله	Thank God	حب	Love
	حلو	sweet	حلو	sweet
	حب	love	الحمدلله	Thank God
	جديد	new	قلب	heart
	قلب	heart	جديد	new
Surprise	ياالله	O God	ياالله	O God
	دفاع	defense	ياشيخ	Oh man
	معقوله	seriously	يمه	O mommy
	يمه	O mommy	معقوله	seriously
	ياشيخ	Oh Man	دفاع	defense
Disgust	يع	yea	عيب	fault
	غبي	stupid	يع	yea
	عيب	fault	غبي	stupid
	قرف	ugh	قرف	ugh
	قلق	worry	سخر	mocked
Trust	امين	Amen	نفسي	myself
	تم	Done	تم	Done
	صدق	believe	امين	Amen
	سر	secret	الصبر	patience
	نفسي	myself	صدق	believe
Anticipation	الله	God	الله	God
	وقت	time	وقت	time
	سلام	peace	شكرا	thank you
	شكرا	thank you	روح	soul
	احس	I feel	نفس	same
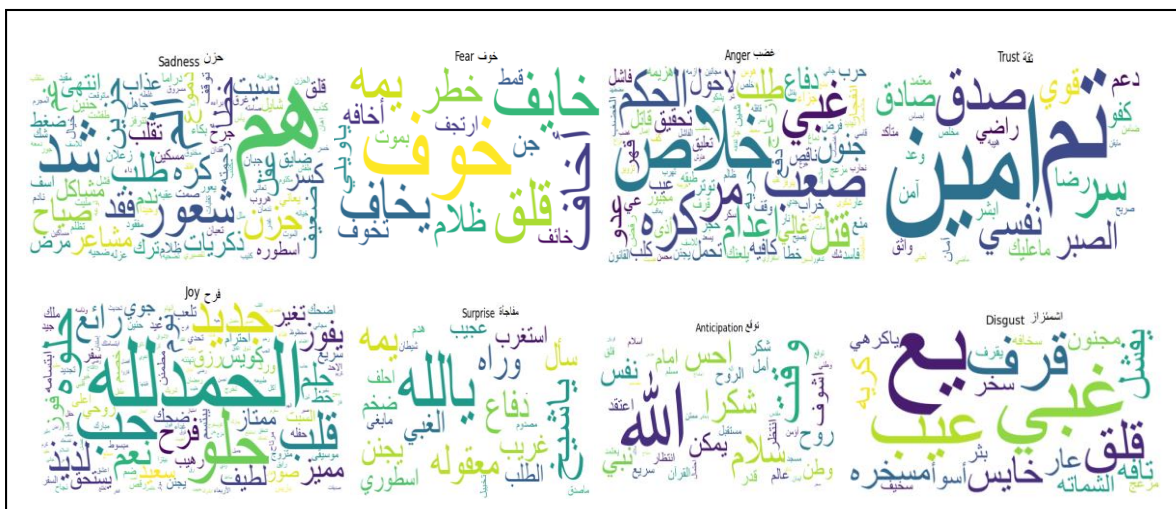
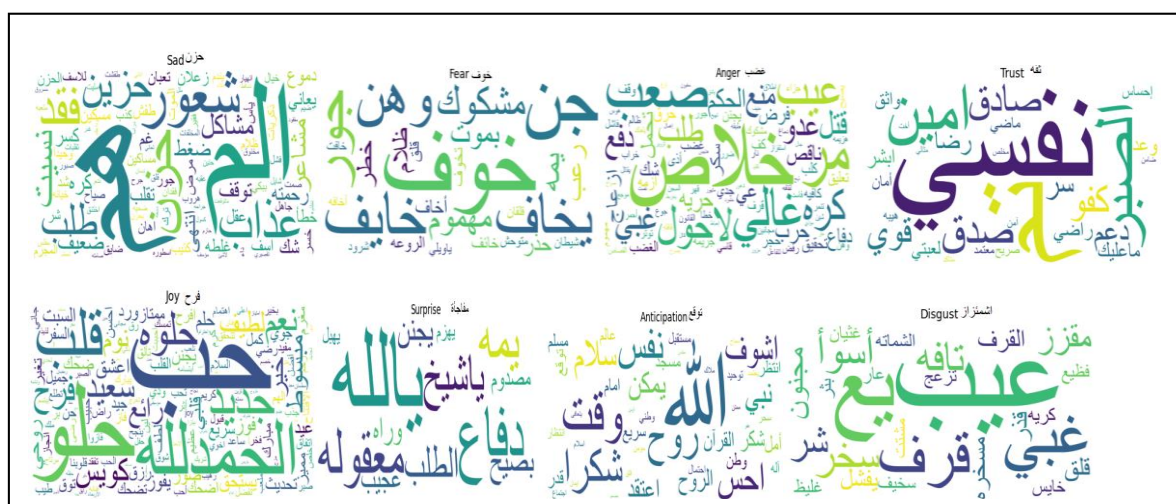**Figure 5.** Word-cloud of Mecca users.

Figure 6. Word-cloud of Riyadh users.

9. Conclusions and Future Scope

In this work, we studied the potential of social media such as Twitter as a platform to depict the emotions and thereby the mental wellbeing of its users. In particular, we studied the contrasting emotions of people residing in a holy city and a typical metropolitan city. To mine emotions from user tweets, we designed a bilingual lexicon (Arabic and English) called AEELex based on two recently proposed emotion lexicons (M. Saad's Arabic Lexicon and the Moodbook English lexicon). The proposed study was an attempt to show that user tweets have the potential to unveil user emotions and thus can help in analyzing user emotions as well as behavior on different platforms. Additionally, emotions are strongly influenced by the surrounding environment. Therefore, in this study, we performed a comparative analysis of tweets posted by users in Riyadh and Mecca (pilgrimage city). Plutchik's eight emotion categories of *Fear*, *Anger*, *Sadness*, *Joy*, *Surprise*, *Disgust*, *Trust*, and *Anticipation* were used for this purpose. To be more vigilant toward the outcomes, the entire analysis was performed on two datasets containing Arabic and English language tweets separately.

The study clearly showed the impact of social and habitual surroundings on user tweets and thereby their emotions. People residing in pilgrimage cities were observed to be more vocal in expressing positive emotions or sentiments. They also conveyed less negativity in comparison to people residing in a non-pilgrimage city. Furthermore, a higher proportion of people residing in a holy city or pilgrimage area expressed emotions in

positive categories such as *Joy* and *Anticipation*, whereas negative emotions were expressed more often by people residing in a typical metropolitan city.

This study could be extended to detect early signs of a psychological disorder such as depression or suicidal tendencies among residents of such contrasting social environments. This social media text-based analysis could be further improved if we also take into account the context of the tweets and also their sarcasm, if present. Furthermore, this analysis could assist in building a machine learning-based model to predict the location (city, state, etc.) of a tweet/post.

Author Contributions: K.A.S. and M.A.W. conceived and designed the experiments; K.A.S. performed the experiments; K.A.S., K.T. and F.S.A. created the bilingual lexicon., K.A.S. and M.A.W. analyzed the data; K.A.S. prepared the first draft of the paper with F.S.A. as Arabic language expert; M.A.W. and K.T. edited the paper; M.A.W. proofread the paper and supervised the overall work. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Institutional Review Board (IRB) at Princess Nourah bin Abdulrahman University, Riyadh, Saudi Arabia.

Informed Consent Statement: Not applicable.

Data Availability Statement: The link to the data repository related to this project can be made available upon request. If required will be made publicly available on GitHub profile (https://bit.ly/3jku4En, accessed on 23 July 2021) under the title of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ulrich, R.S.; Simons, R.F.; Losito, B.D.; Fiorito, E.; Miles, M.A.; Zelson, M. Stress recovery during exposure to natural and urban environments. *J. Environ. Psychol.* **1991**, *11*, 201–230. [CrossRef]
2. Fitri, V.A.; Andreswari, R.; Hasibuan, M.A. Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. *Procedia Comput. Sci.* **2019**, *161*, 765–772. [CrossRef]
3. Kušen, E.; Strembeck, M. Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections. *Online Soc. Netw. Media* **2018**, *5*, 37–50. [CrossRef]
4. Gmi_blogger. Saudi Arabia Social Media Statistics 2020 (Infographics)-GMI Blog. Infographics, Social Media Marketing. 2020. Available online: https://www.globalmediainsight.com/blog/saudi-arabia-social-media-statistics/ (accessed on 28 November 2020).
5. Strapparava, C.; Valitutti, A. WordNet-Affect: An Affective Extension of WordNet. In Proceedings of the LREC, Lisbon, Portugal, 24–30 May 2004.
6. Esuli, A.; Sebastiani, F. Sentiwordnet: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the 5th International Conference on Language Resources and Evaluation, Genoa, Italy, 22–28 May 2006.
7. Basiri, M.E.; Nemati, S.; Abdar, M.; Cambria, E.; Acharrya, U.R. ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. *Future Gener. Comput. Syst.* **2021**, *115*, 279–294. [CrossRef]
8. Adikari, A.; Gamage, G.; de Silva, D.; Mills, N.; Wong, S.-M.J.; Alahakoon, D. A self structuring artificial intelligence framework for deep emotions modelling and analysis on the social web. *Future Gener. Comput. Syst.* **2020**, *116*, 302–315. [CrossRef]
9. Liu, P.; Chen, W.; Ou, G.; Wang, T.; Yang, D.; Lei, K. Sarcasm detection in social media based on imbalanced classification. In *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland; Volume 8485, pp. 459–471.
10. Li, J.; Ni, S.; Kao, H.-Y. Birds of a Feather Rumor Together? Exploring Homogeneity and Conversation Structure in Social Media for Rumor Detection. *IEEE Access* **2020**, *8*, 212865–212875. [CrossRef]
11. Shen, G.; Jia, J.; Nie, L.; Feng, F.; Zhang, C.; Hu, T.; Chua, T.-S.; Zhu, W. Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution. In Proceedings of the IJCAI, Melbourne, Australia, 19–25 August 2017.
12. De Choudhury, M.; Gamon, M.; Counts, S.; Horvitz, E. Predicting Depression via Social Media. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*; Association for the Advancement of Artificial Intelligence: Menlo Park, CA, USA, 2013.
13. Nguyen, T.; Phung, D.; Dao, B.; Venkatesh, S.; Berk, M. Affective and content analysis of online depression communities. *IEEE Trans. Affect. Comput.* **2014**, *5*, 217–226. [CrossRef]

14. Ma, Y.; Cao, Y. Dual Attention based Suicide Risk Detection on Social Media. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2020, Dialian, China, 27 June 2020; pp. 637–640.
15. Plutchik, R. Emotions: A general psychoevolutionary theory. *Approaches Emot.* **1984**, *1984*, 197–219.
16. Wani, M.A.; Agarwal, N.; Jabin, S.; Hussain, S.Z. User emotion analysis in conflicting versus non-conflicting regions using online social networks. *Telemat. Informatics* **2018**, *35*, 2326–2336. [CrossRef]
17. Mohammad, S.M.; Turney, P.D. Crowdsourcing a word-emotion association lexicon. *Comput. Intell.* **2013**, *29*, 436–465. [CrossRef]
18. Li, D.; Rzepka, R.; Ptaszynski, M.; Araki, K. HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. *Inf. Process. Manag.* **2020**, *57*, 102290. [CrossRef]
19. Park, S.B.; Kim, J.; Lee, Y.K.; Ok, C.M. Visualizing theme park visitors' emotions using social media analytics and geospatial analytics. *Tour. Manag.* **2020**, *80*, 104127. [CrossRef]
20. Bagozzi, R.P.; Gopinath, M.; Nyer, P.U. The role of emotions in marketing. *J. Acad. Mark. Sci.* **1999**, *27*, 184–206. [CrossRef]
21. Loureiro, M.L.; Alló, M. Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the U.K. and Spain. *Energy Policy* **2020**, *143*, 111490. [CrossRef]
22. Ghallab, A.; Mohsen, A.; Ali, Y. Arabic Sentiment Analysis: A Systematic Literature Review. *Appl. Comput. Intell. Soft Comput.* **2020**, *2020*, 7403128. [CrossRef]
23. Hawalah, A. Semantic ontology-based approach to enhance arabic text classification. *Big Data Cogn. Comput.* **2019**, *3*, 53. [CrossRef]
24. Abdul-Mageed, M.; Diab, M. SANA: A large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis. In Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, 26–31 May 2014; pp. 1162–1169.
25. Al-Saggaf, Y.; Simmons, P. Social media in Saudi Arabia: Exploring its use during two natural disasters. *Technol. Forecast. Soc. Change* **2015**, *95*, 3–15. [CrossRef]
26. Al-Khatib, A.; El-Beltagy, S.R. Emotional tone detection in Arabic tweets. In Proceedings of the Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Budapest, Hungary, 10 October 2018; Volume 10762, pp. 105–114.
27. Tweepy. Available online: https://www.tweepy.org/ (accessed on 1 April 2021).
28. Klokantech. Bounding Box Tool: Metadata Enrichment for Catalogue Records by Visually Selecting Geographic Coordinates (Latitude/Longitude) for Maps. Available online: https://boundingbox.klokantech.com/ (accessed on 29 October 2020).
29. Gutub, A.; Ghouti, L.; Elarian, Y.; Awaida, S.; Alvi, A. Utilizing Diacritic Marks for Arabic Text Steganography. *Kuwait J. Sci. Eng.* **2010**, *37*, 89–109.
30. Abandah, G.A.; Graves, A.; Al-Shagoor, B.; Arabiyat, A.; Jamour, F.; Al-Taee, M. Automatic diacritization of Arabic text using recurrent neural networks. *Int. J. Doc. Anal. Recognit.* **2015**, *18*, 183–197. [CrossRef]
31. Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; Stede, M. Lexicon-basedmethods for sentiment analysis. *Comput. Linguist.* **2011**, *37*, 267–307. [CrossRef]
32. Jurek, A.; Mulvenna, M.D.; Bi, Y. Improved lexicon-based sentiment analysis for social media analytics. *Secur. Inform.* **2015**, *4*, 9. [CrossRef]
33. Turney, P.D.; Littman, M.L. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inf. Syst.* **2003**, *21*, 315–346. [CrossRef]
34. Tong, R. An operational system for detecting and tracking opinions in on-line discussions. In Proceedings of the Working Notes of the SIGIR Workshop on Operational Text Classification, New York, NY, USA, 13 September 2001; pp. 1–6.
35. Saad, M. *Mining Documents and Sentiments in Cross-Lingual Context*; Université de Lorraine: Metz, France, 15 February 2015.
36. Koenig, H.G.; Al Zaben, F.; Sehlo, M.G.; Khalifa, D.A.; Al Ahwal, M.S.; Qureshi, N.A.; Al-Habeeb, A.A. Mental Health Care in Saudi Arabia: Past, Present and Future. *Open J. Psychiatry* **2014**, *4*, 113–130. [CrossRef]
37. Plutchik, R. *The Emotions*; University Press of America: Lanham, MD, USA, 1991.