Exploring Differences in the Impact of Users' Traces on Arabic and English Facebook Search

Ismail Badache

Exploring Differences in the Impact of Users' Traces on Arabic and English Facebook Search

Ismail Badache
Aix Marseille Univ, Université de Toulon, CNRS, LIS
Marseille, France
Ismail.Badache@lis-lab.fr

ABSTRACT

This paper proposes an approach on Facebook search in Arabic and English, which exploits several users' traces (e.g. comment, share, reactions) left on Facebook posts to estimate their social importance. Our goal is to show how these social traces (signals) can play a vital role in improving Arabic and English Facebook search. Firstly, we identify polarities (positive or negative) carried by the textual signals (e.g. comments) and non-textual ones (e.g. the reactions love and sad) for a given Facebook posts. Therefore, the polarity of each comment expressed in Arabic or in English on a given Facebook post, is estimated on the basis of a neural sentiment model. Secondly, we group signals according to their complementarity using attributes (features) selection algorithms. Thirdly, we apply learning to rank (LTR) algorithms to re-rank Facebook search results based on the selected groups of signals. Finally, experiments are carried out on 13,500 Facebook posts, collected from 45 topics, for each of the two languages. Experiments results reveal that Random Forests was the most effective LTR approach for this task, and for the both languages. However, the best appropriate features selection algorithms are *ReliefFAttributeEval* and *InfoGainAttributeEval* for Arabic and English Facebook search task, respectively.

KEYWORDS

Facebook Search, Sentiment Analysis, User Generated Content

ACM Reference Format:
Ismail Badache. 2019. Exploring Differences in the Impact of Users' Traces on Arabic and English Facebook Search. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI '19), October 14–17, 2019, Thessaloniki, Greece.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3350546.3352522

1 INTRODUCTION

Social media has largely contributed to the launch of the so-called *Arab Spring*. Since then, the penetration of social media has grown steadily. The number of Facebook users in the Arab world is estimated at 164 million among 2.28 billion users[1]. English language is the first used language on Facebook, 1.1 billion Facebook users speak English. That leaves the other half using the platform in over 100 other languages. The second most popular language on Facebook is Spanish, at 310 million, followed by Indonesian at 170 million, and Arabic and Portuguese, at 150 million each. So, there are about 14 million Arab users who do not speak Arabic on Facebook[2]. This movement reflects the democratization of the ways of production and interaction in the Web (user-generated content) thanks to new technologies. Among these ways increasingly accessible to a wide audience include social networks, blogs, microblogs, etc. User-Generated Content (UGC) refers to a set of data (e.g. comments, posts, reactions, signals) whose content is primarily either produced or directly influenced by end users.

The main task in information retrieval (IR) is to find a set of relevant documents to a specific information need (query). For this, effective approaches have existed for many years that exploit two classes of features to rank documents responding to a given query. The first class, the most used one, is query-dependent, which includes features corresponding to particular statistics of query terms such as term frequency, and term distribution within a document or in the collection of documents. The second class corresponds to query-independent features, which measure the a priori importance of the document. For example, the number of backlinks [28], page length and URL form [43], PageRank [16], document authors [30] and social signals [8, 9].

This paper explores differences in the impact of users' traces (like, share, positive comment, negative comment, love, haha, angry, wow and sad) on the effectiveness of the relevance ranking of Arabic and English Facebook search. In order to design our Arabic and English IR approach, fundamental tasks are carried out. First, we identify the polarity for each comment left on a given post using a neural sentiment analysis in Arabic and English languages. Then, we use features (attributes) selection algorithms to identify the most fruitful features (users' traces) for Arabic and English social IR task. Finally, we evaluate the impact of these features on the relevance of Arabic and English Facebook search results. More specifically, we try to select and compare the most effective features and combine them with Learning-To-Rank (LTR) approaches to improve Arabic and English IR on Facebook. The main contributions discussed in this paper are twofold:

(C1). Evaluate the impact of social features (users' traces and comment sentiment) on Arabic and English Facebook search. We try to answer the following research questions: a) What are the best social features suitable for this task? ; b) What is the impact of these features on the performance of Facebook's search engine in both Arabic and English?

[1] https://arabiawithclass.com/164-million-active-facebook-users-in-the-arab-world-study-shows/

[2] https://blog.hootsuite.com/facebook-statistics/

(C2). Build Arabic and English datasets (documents, topics, qrels) from Facebook. These datasets are useful to evaluate social IR systems in Arabic and English languages. User studies are conducted to collect relevance judgments.
(C3). Objectively compare the impact of various social features on Arabic and English Facebook search?

The paper is organized as follows. Section 2 presents some background and related work. Section 3 details our social IR approach. Section 4 presents our test datasets. The experimental evaluation is presented in section 5. Finally, Section 6 concludes this paper with some perspectives.

2 BACKGROUND AND RELATED WORKS

This section presents an overview of the Social Information Retrieval (SIR) and their major components related to our work. Beginning with a presentation of different types of UGCs, description and interrelationships of the English and Arabic sentiment analysis and our SIR approach. Then a focused overview of SIR approaches exploiting users' traces and social networks is presented.

2.1 User Generated Content

User Generated Content is often linked to a specific social network with different operating rules (see table 1). The popularity of UGCs, especially in the context of social media, has given rise to many new problems in IR [15]. More specifically, how to exploit these social contents in favor of IR is an open question. In the following, we present an overview on the search for social information.

Table 1: List of different types of UGCs (social signals)

Type	Example	Social Networks
Vote	Like, +1	Facebook, LinkedIn, Google+, StumbleUpon
Message	Tweet, Post	Facebook, Google+, LinkedIn, Twitter
Share	Share, Re-tweet	Google+, Twitter, Buffer, Facebook, LinkedIn
Tag	Bookmark, Pin	Delicious, Pinterest, Diigo, Digg
Comment	Comment, Reply	Facebook, Google+, LinkedIn, Twitter
Emotion	Love, Haha, Wow	Facebook
Event Reaction	Thankful (Mother's day)	Facebook
Relation	Followers, Friends	Facebook, Twitter

2.2 English Sentiment Analysis

English sentiment analysis has been the subject of much previous research. Supervised and unsupervised approaches both propose their solutions. Thus, some unsupervised approaches are based on lexicons, such as the approach developed by [42], or corpus-based methods, such as in [33]. *Pang et al.* [37] proposed supervised approaches, that perceive the task of sentiment analysis as a classification task and therefore use methods such as SVM (Support Vector Machines) or Bayesian networks. Other recent studies are based on RNN (Recursive Neural Network), such as in [41]. In our work, sentiment analysis is only a part of English social IR process, we used the approach proposed in [38] to measure comment polarity and consider it as an additional social feature. The approach proposed by Radford *et al.* [38] is described in the section 3.1.

2.3 Arabic Sentiment Analysis

Arabic sentiment analysis is useful for quantifying the polarity of Arabic textual UGCs such as comments and tweets. However, to the best of our knowledge, just a few works have been done on sentiment analysis in Arabic language. This can be explained by the lack of standard datasets. Farra *et al.* [24] proposed a linguistic method based on a set of patterns to extract the polarities from a financial document. Abdulla *et al.* [2] have manually built a lexicon containing 4815 words. Their system calculates the number of positive and negative words in a text in order to generate its global polarity. Al-Kabi *et al.* [4] have set up a tool that determines the subjectivity, the polarity of an opinion and its intensity. They used two general lexicons and sixteen specific lexicons. Abdulla *et al.* [3] proposed a statistical approach to detect subjectivity and polarity in social networks using morphological attributes. Bayoudhi *et al.* [14] compared three classifiers: SVM, Naive Bayes and a simple neural network. El-Halees *et al.* [22] proposed a hybrid system for the opinions analysis in Arabic language. He proposed a sequential hierarchy of combined classifications. Ibrahim *et al.* [26] used a lexicon of 5244 adjectives, a lexicon of 3296 idioms to improve sentences classification with using SVM. Refaee and Rieser [40] applied a hybrid approach for predicting the intensity of polarity in tweets. They used logistic regression specifically to predict initial scores that are adjusted by applying rules extracted from a polarity lexicon. Other recent works apply deep learning techniques for opinion analysis [13, 20]. Barhoumi *et al.* [13] used continuous representations of documents combined with a MultiLayer Perceptron (MLP) while Dahou *et al.* [20] used CNN (Convolutional Neural Network). Barhoumi *et al.* [12] illustrated a relevant comparison between several systems of Arabic sentiments detection, experienced in the Large-scale Arabic Book Review dataset (LABR)[3]. They showed that the best results were obtained by Dahou *et al.* [20] using CNN (77.39% of accuracy). The second best system is that of ElSahar and El-Beltagy [23], they have built a large Arabic lexicon multi-domains for sentiment analysis. The reviews was collected from various websites (e.g. hotels[4], movies[5], products[6] and restaurants[7]). We recall that our goal in this paper is exploiting social features to improve Arabic Facebook search. For this, we used the approach proposed in [20] (see section 3.1) to measure comment polarity and consider it as an additional relevance factor. We therefore used the approach proposed by Dahou *et al.* [20] that we describe in the section 3.1.

2.4 Social Information Retrieval

Social Information Retrieval (SIR) has extended traditional IR with different social features in order to satisfy social motivations behind the user's information needs. In 2012, Jaime Teevan[8], a researcher at Microsoft, defines social IR as follows: "*Social search is an emerging research area that explores how social interactions and social data can enhance existing information-seeking experiences, as well as enable new information retrieval scenarios. This session will showcase different models of social search, including 1) the use of social data to augment search, 2) social data as new information to be searched, and 3) social interaction and collaboration as part of the search process.*"

[3] https://github.com/mohamedadaly/LABR
[4] https://www.tripadvisor.com/
[5] https://www.elcinema.com/
[6] https://uae.souq.com/ae-ar/
[7] http://www.qaym.com/
[8] https://www.microsoft.com/en-us/research/video/social-search-panel/

In addition, novel social approaches are coming into play to meet the new information needs and IR requirements initiated by social practices on the Web. We consider the social IR according to 3 axes: 1) the first axis concerns the search for information of a social nature. It's about finding social information that responds to the user. We distinguish, for example, the information retrieval in blogs, microblogs [21, 29] and the search for conversations [31]; 2) the second concerns the exploitation of social contents to improve IR, in which social information is used to improve the information retrieval process, for example, tags in folksonomies have been found useful for improving web search and personalized search [18], re-ranking of search results [9, 17] and 3) the third paradigm concerns the search for information carried out by several people, collaborative search [44].

Our work concerns both the first and second axes mentioned by Jaime Teevan, we propose an approach to improve Facebook search using its UGCs. While considerable work has been done in the context of social IR in English language, there is still a lack of studies that would analyze the impact of users' traces on Facebook search in Arabic language. The most related works to ours include [6, 7, 9, 17, 19, 36]. These works focus on the exploitation of social features to improve IR in English on the Web and on social networks. The approach we propose in this paper is in the same vein as these works, i.e exploiting social features around Facebook posts (documents) to improve the ranking of search results. However, our work differs from the state of the art in the following points. First, in addition to English, our approach is to search for information in Arabic language on Facebook. A sentiment analysis of the comments left by the users on a given publication is necessary. Next, we use Learning To Rank algorithms combined with features selection techniques. More specifically, we estimate the social importance of a Facebook post by exploiting these social traces (like, share, polarity of comment: positive or negative, love, haha, angry, wow and sad) to improve the effectiveness of Arabic and English search in Facebook.

3 EXPLOITING USERS' TRACES TO IMPROVE BILINGUAL FACEBOOK SEARCH ENGINE

Our approach is based both on the classical traces (e.g. the frequencies of the signals "like" and "share", etc.) and on the emotional traces (e.g. the frequencies of the reactions "love" and "sad", etc.) as well as on the sentiment analysis of the comments expressed on each Facebook post (document). We note that Facebooks reactions allow users to express more nuanced emotions compared to classical signals. The goal of our approach is to improve the relevance of the results returned by the Facebook search engine in Arabic and English languages by exploiting all these Facebook traces (or signals). They are considered as a priori knowledge to be taken into account in the Arabic and English Facebook search process.

3.1 Social Traces-Based Search Process

Three main steps are required: 1) extracting features and estimating sentiments for all comments; 2) selecting the best features for SIR task; and 3) combining LTR algorithms with selection techniques. The figure 1 illustrates our adopted Learning To Rank (LTR) process.

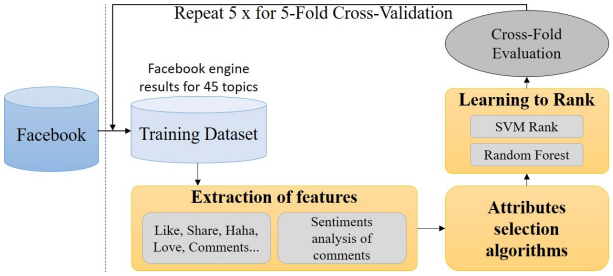

Figure 1: LTR process for bilingual Facebook search

English Sentiment Analysis. The sentiment of the comment expressed in English is estimated using *SentiNeuron*[9], an unsupervised model proposed by Radford *et al.* [38] to detect sentiment signals in reviews. The model consisted of a single layer multiplicative long short-term memory (mLSTM) cell and when trained for sentiment analysis it achieved state of the art on the movie review dataset[10]. They also found a unit in the mLSTM that directly corresponds to the sentiment of the output.

Arabic Sentiment Analyzer. The sentiment of comment is estimated using the model proposed by Dahou *et al.* [20] whose implementation is publicly available[11]. Dahou *et al.* [20] proposed a CNN approach to identify the polarity of Arabic comments. When considering the semantics of words, it has been shown that neural word embedding captures semantic similarities between the words [32]. Such distributed representations of words in a dense vector space are learned efficiently on large collections. Therefore, Dahou *et al.* [20] investigated different neural word embedding architectures using a corpus of 3.4 billion words chosen from a collected web-crawled corpus of 10 billion words. Then, the CNN was trained on top of the pre-trained word embeddings to classify the sentiments without considering aspect-level (topic on which the sentiment is concerned). They trained the model word2vec on web pages [32] using Skip-gram (SKIP-G) and Continuous Bag Of Words (CBOW) methods of constructing the training data for the neural network. Their experiments results showed that CBOW is more efficient and their architecture outperforms existing methods on several publicly available datasets presented in [1, 5, 23, 35, 39].

Selection of the Best Relevance Features. The difficulty in our approach lies on the selection of features groups. It is impossible to judge properly and rigorously the complementarity of the social features, and to identify the best combinations without testing all possible combinations. In this step, we relied on features selection techniques to determine the best features groups that can be considered into the LTR of the IR process.

Combining LTR Algorithms with Selection Techniques. In this step, we wondered whether the features selection really improves the results of a Arabic and English Facebook search task. We relied on the work of Hall and Holmes [25]. We studied the effectiveness of some features selection techniques by confronting them with

[9] https://github.com/openai/generating-reviews-discovering-sentiment
[10] https://www.cs.cornell.edu/people/pabo/movie-review-data/
[11] https://pan.baidu.com/s/1eS2mxCe#list/path=%2F

LTR algorithms. Since the performance of social features differs from one LTR algorithm to another, we identified the best features selection techniques to find the best performing features according to the LTR algorithms.

4 DATASETS AND RELEVANCE JUDGMENTS

To the best of our knowledge, there is no standard Arabic or English Facebook datasets containing posts, users' traces, topics and qrels to evaluate the effectiveness of Arabic and English IR on Facebook. Therefore, we collected 13,500 posts with their users' traces, for each language (Arabic and English), extracted from Facebook via its API and also using *parsing*, between 16 and 28 January 2018. These data were collected via the Facebook search engine for 45 topics (Table 2 shows an example of these topics) that we have defined (nature of topics: Politic: 42%, Sport: 24%, Art: 18%, Leisure: 11%, Other: 5%). Tables 3 and 4 give some statistics about our Arabic and English datasets, respectively. It presents the 10 features we considered for estimating the relevance of Facebook posts for a given topic. The nature of the features from f_1 to f_{10} is a simple count, for example the feature f_1 and f_4 represent the number of "Like" and emotional reaction "Sad", generated on the document. Concerning the last two features f_9 and f_{10}, they represent the number of opinions expressed on the document according to their polarity (positive or negative), respectively. These two features are calculated based on a specific sentiment model presented in the previous sections 3.1 and 3.1.

Table 2: Examples of Arabic and English topics (queries)

Arabic topic	English topic (translation)
الطفل السوري عمران	The Syrian child Omran
اضراب جامعة بيرزيت	The strike of Birzeit University
قتل السفير الروسي	The assassination of the Russian ambassador
رقصة الزومبا	The Zumba Dance

Table 3: Arabic Facebook Data Statistics

Posts (documents) & Topics		13,500 Arabic documents	45 Arabic topics			
f_i	Feature	Description	SUM	MIN	MAX	AVG
f_1	Like	*#Like* on the document	2031958	0	32025	151
f_2	Share	*#Share* on the document	2329934	0	16781	173
f_3	Comment	*#Comment* on the document	2717589	0	24306	201
f_4	Sad	*#Sad* on the document	63970	0	80	5
f_5	Angry	*#Angry* on the document	95752	0	119	7
f_6	Love	*#Love* on the document	397679	0	496	29
f_7	Haha	*#Haha* on the document	246715	0	308	18
f_8	Wow	*#Wow* on the document	171234	0	213	13
f_9	Positive Comment	*#PositiveComment* on the document	1527546	0	13750	113
f_{10}	Negative Comment	*#NegativeComment* on the document	1134831	0	10063	84

Table 4: English Facebook Data Statistics

Posts (documents) & Topics		13,500 English documents	45 English topics			
f_i	Feature	Description	SUM	MIN	MAX	AVG
f_1	Like	*#Like* on the document	5517756	0	86963	409
f_2	Share	*#Share* on the document	6326907	0	45569	469
f_3	Comment	*#Comment* on the document	7379579	0	66003	547
f_4	Sad	*#Sad* on the document	173710	0	217	13
f_5	Angry	*#Angry* on the document	260013	0	323	19
f_6	Love	*#Love* on the document	1079892	0	1347	80
f_7	Haha	*#Haha* on the document	669951	0	836	50
f_8	Wow	*#Wow* on the document	464984	0	57	34
f_9	Positive Comment	*#PositiveComment* on the document	4148032	0	37338	307
f_{10}	Negative Comment	*#NegativeComment* on the document	3081619	0	27326	228

To obtain the relevance judgments for a given topic: a) 6 users were asked to assess the first 300 documents returned for a given topic using a 3-point relevance scale (irrelevant, somewhat relevant and relevant). Each topic is judged by 3 users. To avoid any bias, none of the social features were displayed with the documents, but all textual content, images or video (according to the Facebook post) are displayed to facilitate the task of judgment. We computed the agreement degree between assessors for each topic using Kappa Cohen measure k. The k ranges from 0.45 to 0.90 and from 0.50 to 0.95 for Arabic and English topics, respectively. The average measure of agreement between the assessors is 75% (strong agreement) and 80% (strong agreement) for Arabic and English topics, respectively.

5 EXPERIMENTAL INVESTIGATION

In order to make a fair and objective comparison between the impact of the users' traces on the Arabic and the English Facebook search engine, we conducted a series of experiments on the two datasets, in Arabic and in English. We compared our approach which takes into account social features, with the baseline formed by only a textual model without considering any social features. Then, we compared the results of Arabic SIR approach with those of English SIR approach. Our main goal in these experiments is to evaluate and compare the impact of the proposed social features on bilingual Facebook search : in Arabic and in English.

5.1 Identification of the Most Effective Features

In order to understand the real impact of the different social features, we evaluated the impact of each of them by using features selection techniques. The goal is to determine the best features to exploit in the LTR algorithms. Features selection techniques aim to identify and remove the maximum amount of unnecessary, redundant and irrelevant information upstream of a learning-based process [25]. They also make it possible to automatically select the subsets of features for obtaining the best results. We used Weka[12] for these experiments, it is a powerful open-source Java-based learning tool that brings together a large number of learning algorithms and features selection techniques.

We proceeded as follows: we identified relevant and irrelevant documents (Arabic and English posts) according to the "qrels", for the top 300 documents for each topic (45 Arabic and English topics) returned by the default Facebook search engine. The resulting set contains 13,500 documents for each language including: 2971 and 5193 relevant Arabic and English documents, respectively; 10529 and 8307 irrelevant Arabic and English documents, respectively. We observed that these collections have an unbalanced relevance classes distribution. This occurs when there are many more elements in one class than in the other class of a training collection. In this case, a LTR algorithm usually tends to predict samples from the majority class and completely ignore the minority class. For this reason, we applied an approach to sub-sampling (reducing the number of samples that have the majority class) to generate a balanced collections composed of: 2971 and 5193 relevant Arabic and English documents, respectively; 2971 and 5193 irrelevant Arabic and English documents, respectively. Irrelevant documents for this study were selected randomly. Finally, we applied the selection algorithms on the two sets obtained, for 5-folds cross-validation.

Features selection algorithms consist in assigning a score to each feature according to its significance for the relevance class (relevant and irrelevant). These algorithms return importance ranking of the features according to the number of times that a given feature has

[12] http://www.cs.waikato.ac.nz/ml

been selected by the algorithm in the cross-validation. We note that we used for each algorithm the default setting provided by Weka.

Tables 5 and 6 present the selected features by the 12 features selection algorithms for the Arabic and English Facebook datasets, respectively. A feature selected by the algorithm is a feature designated by a "+" and an unselected feature is designated by a "-". The results are discussed in the following.

For Arabic Facebook dataset (see Table 5), we remark that the features f_{10}: *Negative Comment*, f_9: *Positive Comment*, f_1: *Like*, f_2: *Share* and f_6: *Love* are the most selected and height ranked compared to other features. The features f_4: *Sad*, f_5: *Angry* are moderately favored by the features selection algorithms, except algorithms *CfsSubsetEval*, *WrapperSubsetEval* and *GainRatioAttributeEval* that did not selected them. The features f_7: *Haha*, f_8: *Wow* are only selected by 7 and 6 algorithms, respectively. Finally, the weakest and most disadvantaged feature is the f_3: *Comment*, it is only selected by 5 out of 12 algorithms.

Table 5: The selected features by the selection algorithms (Arabic Facebook Dataset)

Algorithm	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}
CfsSubsetEval	+	+	-	-	-	+	-	+	+	+
WrapperSubsetEval	+	+	-	-	-	+	-	-	+	+
ConsistencySubsetEval	+	+	+	+	+	+	+	+	+	+
FilteredSubsetEval	+	+	-	+	+	+	-	-	+	+
ChiSquaredAttributeEval	+	+	+	+	+	+	+	+	+	+
FilteredAttributeEval	+	+	+	+	+	+	+	-	+	+
GainRatioAttributeEval	+	+	-	-	-	+	-	+	+	+
InfoGainAttributeEval	+	+	+	+	+	+	+	+	+	+
OneRAttributeEval	+	+	+	+	+	+	+	-	+	+
ReliefFAttributeEval	+	+	-	+	+	+	+	+	+	+
SVMAttributeEval	+	+	-	+	+	+	+	-	+	+
SymetricalUncertEval	+	+	-	+	+	+	-	-	+	+
Total	12	12	5	9	9	12	7	6	12	12

For English Facebook dataset (see Table 6), we remark that the features f_{10}: *Negative Comment*, f_9: *Positive Comment*, f_1: *Like*, f_2: *Share* and f_4: *Sad* are the most selected and height ranked compared to other features. The features f_5: *Angry* and f_8: *Wow* are moderately favored by the features selection algorithms, except the algorithms *CfsSubsetEval*, *GainRatioAttributeEval* that did not selected f_5 and the algorithms *FilteredSubsetEval*, *OneRAttributeEval*, *SymetricalUncertEval* that did not selected f_8. The features f_3: *Comment*, f_6: *Love* are only selected by 7 and 6 algorithms, respectively. Finally, the weakest and most disadvantaged feature is the f_7: *Haha*, it is only selected by 5 out of 12 algorithms.

Table 6: The selected features by the selection algorithms (English Facebook Dataset)

Algorithm	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}
CfsSubsetEval	+	-	+	+	-	+	+	+	+	+
WrapperSubsetEval	+	+	-	+	+	+	-	+	+	+
ConsistencySubsetEval	+	+	+	+	+	-	+	+	+	+
FilteredSubsetEval	+	+	+	+	+	-	-	-	+	+
ChiSquaredAttributeEval	+	+	+	+	+	-	+	+	+	+
FilteredAttributeEval	+	+	+	+	+	+	-	+	+	+
GainRatioAttributeEval	+	+	-	+	-	+	-	+	+	+
InfoGainAttributeEval	+	+	-	+	+	-	-	+	+	+
OneRAttributeEval	+	+	+	+	+	+	-	-	+	+
ReliefFAttributeEval	+	+	+	+	+	-	+	+	+	+
SVMAttributeEval	+	+	-	+	+	+	+	+	+	+
SymetricalUncertEval	+	+	-	+	+	-	-	-	+	+
Total	12	11	7	12	10	6	5	9	12	12

5.2 Social Features-Based Learning to Rank

Other experiments were carried out exploiting these social features in supervised approaches based on LTR models. We used the instances (Facebook posts) of the 45 topics as training sets. Then we used two LTR algorithms. This choice is explained by the fact that they often showed their effectiveness in IR: RankSVM [27] and Random Forests [34]. Regarding RankSVM, we use the implementation[13] with its default settings proposed by Joachims [27]. While for Random Forests, we used Weka's implementation[14]. We have set the option "*max depth*" to 0 (unlimited) and the number of trees to 100. The input of each algorithm is a vector of features (see tables 3 and 4), that is all the features or only the features selected by a given selection algorithm. LTR algorithms predict the relevancy ranking of search results. Finally, we applied a cross-validation for 5-folds.

The question at this stage is related to the specification of the input features vector for the Learning To Rank algorithms, either we take all the features, or we keep only those selected by the features selection techniques. In this case, with which LTR algorithms these will be combined.

In order to take into account the selected social features in LTR models, we have carried out several experiments to identify the best features selection techniques allowing to find the most effective features according to the LTR techniques. Based on this study, for Arabic Facebook dataset, we found the following best pairs of LTR algorithms and the features selection techniques: a) Features selected by *CfsSubsetEval* (CFS) and *WrapperSubsetEval* (WRP) are learned using RankSVM and Random Forests; b) Features selected by *ReliefFAttributeEval* (RLF) are learned using Random Forests; and c) Features selected by *SVMAttributeEval* (SVM) are learned using RankSVM. The same applies to the English Facebook dataset, with the exception of the features selected by *InfoGainAttributeEval* (IGA), which is learned using Random Forests. We recall that features selection algorithms have highlighted the following mains sets of features:

Table 7: Selected features sets (Arabic Facebook Dataset)

Selection algorithms	Selected features
CfsSubsetEval (CFS)	$f_1, f_2, f_6, f_8, f_9, f_{10}$
WrapperSubsetEval (WRP)	$f_1, f_2, f_6, f_9, f_{10}$
SVMAttributeEval (SVM)	$f_1, f_2, f_4, f_5, f_6, f_7, f_9, f_{10}$
ReliefFAttributeEval (RLF)	$f_1, f_2, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}$

Table 8: Selected features sets (English Facebook Dataset)

Selection algorithms	Selected features
CfsSubsetEval (CFS)	$f_1, f_3, f_4, f_6, f_7, f_8, f_9, f_{10}$
WrapperSubsetEval (WRP)	$f_1, f_2, f_4, f_5, f_6, f_8, f_9, f_{10}$
SVMAttributeEval (SVM)	$f_1, f_2, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}$
InfoGainAttributeEval (IGA)	$f_1, f_2, f_4, f_5, f_8, f_9, f_{10}$

In order to check the significance of the results compared to Facebook (baseline), we conducted the Student's t-test. We attached

[13] http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html
[14] http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/RandomForest.html

* (strong significance) and ** (very strong significance) to the results in tables 9 and 10 when *p-value* < *0.05* and *p-value* < *0.01*, respectively. The results are discussed in the following.

Tables 9 and 10 compare the different configurations of our bilingual SIR approach in terms of precision@k ($k \in \{5, 10\}$), nDCG and MAP, obtained by RankSVM and Random Forests using the features emerged from the study using features selection techniques: CFS, WRP, SVM and RLF. We notice globally, with the consideration of social features, the results obtained are significantly better than those obtained by Facebook model (baseline). The results are discussed below for each LTR algorithm.

5.3 Discussion: Arabic Facebook Dataset

Results obtained by RankSVM. According to the table 9, the results obtained by RankSVM using the selection algorithm *SVMAttributeEval* (SVM), where only the two features f_3 and f_8 were not selected, are better than those obtained using (CFS, WRP or all features). We recorded improvement rates of 57% and 65% in terms of nDCG and MAP, respectively, compared to the baseline model. Using CFS which selects only 6 features f_1, f_2, f_6, f_8, f_9, f_{10}, and WRP which selects even fewer features f_1, f_2, f_6, f_9, f_{10}, the results fall with rates of −25% and −32% in terms of nDCG, respectively. Consequently, the unselected features f_4, f_5, f_7 and f_8 are fruitful for RankSVM. In addition, with the selection of all features, RankSVM achieves better results than those obtained with CFS and WRP when certain features are ignored. Indeed, some topics such as (translation Arabic to English: *the Syrian child Omran*) and (translation Arabic to English: *blockade of Gaza*) recorded the highest precision when the features f_4: *Sad* and f_5: *Angry* are taken into account (with 0.8957 and 0.9324 in terms of P@10, respectively). The features f_8: *Wow* and f_7: *Haha* are more effective with topics that represent weird, exciting, or funny information. Finally, even if the RankSVM algorithm is expensive in terms of execution time, it remains favorable to obtain significant results. We noticed that RankSVM combined with the selection algorithm (SVM) obtained the second best result after the results obtained by Random Forests combined with the *ReliefFAttributeEval* (RLF) selection algorithm.

Results obtained by Random Forests. According to the table 9, the results confirm that the Random Forests decision tree is the most appropriate model when combined with the selection algorithm *ReliefFAttributeEval* (RLF), it takes into account all the features, except for the feature f_3: *Comment*, more efficiently than the other configurations (improvement rates of 80% and 92% in terms of nDCG and MAP compared to baseline model, respectively). The improvement rates compared to the baseline model using CFS and WRP are relatively low (18% and 13% in terms of nDCG, respectively). We also note that Random Forests (combined with the RLF selection algorithm) exceeds the best RankSVM configuration (combined with the SVM selection algorithm) with a rate of 15% and 16% in terms of nDCG and MAP, respectively. In addition, the improvements are also highly significant for the configuration taking all the features with Random Forests (ranked 3^{rd} after Random Forest with RLF and RankSVM with SVM).

5.4 Discussion: English Facebook Dataset

Results obtained by RankSVM. According to the table 10, the results obtained by RankSVM using the selection algorithm *SVMAttributeEval* (SVM), where only the feature f_3 was not selected, are better than those obtained using (CFS, WRP or all features). We recorded improvement rates of 27% and 31% in terms of nDCG and MAP, respectively, compared to the baseline model. Using CFS which selects 8 features f_1, f_3, f_4, f_6, f_7, f_8, f_9, f_{10}, and WRP which selects the same number of features as CFS (8 features) but differently as follows: f_1, f_2, f_4, f_5, f_6, f_8, f_9, f_{10}, the results fall with rates of −5% in terms of nDCG compared to SVM. Consequently, the unselected features f_5 and f_7 are fruitful for RankSVM. In addition, with the selection of all features, RankSVM achieves worse results than those obtained with the all other configurations. Indeed, in English, some topics such as *blockade of Gaza* recorded the highest precision when the features f_4: *Sad*, f_5: *Angry* and f_8: *Wow* are taken into account. As in Arabic search, the feature f_7: *Haha* is more effective with topics that represent funny information. Finally, RankSVM obtain the fourth best significant results when combined with the selection algorithm (SVM).

Results obtained by Random Forests. According to the table 10, the results confirm again that the Random Forests decision tree is the most appropriate model. However, in English search, the best results are obtained when Random Forest is combined with the selection algorithm *InfoGainAttributeEval* (IGA), it takes into account the most important features, except for the feature f_3: *Comment*, more efficiently than the other configurations (improvement rates of 79% and 91% in terms of nDCG and MAP compared to baseline model, respectively). The improvement rates compared to the baseline model using CFS and WRP are relatively low. We also note that Random Forests (combined with the IGA selection algorithm) exceeds the best RankSVM configuration (combined with the SVM selection algorithm) with a rate of 41% and 46% in terms of nDCG and MAP, respectively. In addition, the improvements are also highly significant when the feature f_3: *Comment* is ignored. Configurations considering all the features with Random Forests or RankSVM are the lowest ranked, 7^{th} and 8^{th}, respectively.

5.5 Discussion: English Vs Arabic on Facebook

After these series of experiments using two Facebook datasets, in Arabic and in English on a similar information need (query i.e. topics), we identified these two findings:

- An Arabic or English speaking profile on Facebook is directly influenced by the events of his community, his region, his culture. Consequently, signals generated by these profiles are also influenced by the culture of their creator (language, beliefs, interests, region, etc.).
- For example for the two queries *the Syrian child Omran* and *blockade of Gaza*: for an Arabic profile, a user who lives in an Arabic environment, close and aware of the information carried by these two queries reacts specifically by a sadness or anger through the Facebook reactions "Sad" and "Angry". As a result, Arabic language posts that may involve this type of user are likely to have much more sadness and anger types reactions as well as comments with negative polarity. While

Table 9: LTR results of P@{5, 10}, nDCG et MAP (Arabic Facebook Dataset)

IR Model		P@5	P@10	nDCG	MAP
Facebook (baseline model)		0.1911	0.1721	0.2513	0.1002
LTR Algorithms	Selection Algorithms	P@5	P@10	nDCG	MAP
RankSVM	CfsSubsetEval (CFS)	0.2133*	0.1944*	0.2955*	0.1204*
	WrapperSubsetEval (WRP)	0.1992	0.1802	0.2674	0.1076
	SVMAttributeEval (SVM)	0.2627**	0.2441**	0.3939**	0.1654**
	All features	0.2254*	0.2066*	0.3196*	0.1314*
Random Forests	CfsSubsetEval (CFS)	0.2395*	0.2046*	0.2955*	0.1149*
	WrapperSubsetEval (WRP)	0.2072	0.1883	0.2834	0.1149
	ReliefFAttributeEval (RLF)	0.2920**	0.2735**	0.4522**	0.1921**
	All features	0.2526**	0.2340**	0.3738**	0.1563**

Table 10: LTR results of P@{5, 10}, nDCG et MAP (English Facebook Dataset)

IR Model		P@5	P@10	nDCG	MAP
Facebook (baseline model)		0.3340	0.3008	0.4392	0.1751
LTR Algorithms	Selection Algorithms	P@5	P@10	nDCG	MAP
RankSVM	CfsSubsetEval (CFS)	0.3728*	0.3398*	0.5165*	0.2008*
	WrapperSubsetEval (WRP)	0.4186*	0.3576*	0.5165*	0.2104*
	SVMAttributeEval (SVM)	0.3940*	0.3611*	0.5586*	0.2297*
	All features	0.3482	0.3150	0.4674	0.1881
Random Forests	CfsSubsetEval (CFS)	0.4415**	0.4090**	0.6533**	0.2732**
	WrapperSubsetEval (WRP)	0.4591**	0.4266**	0.6885**	0.2891**
	InfoGainAttributeEval (IGA)	0.5104**	0.4780**	0.7904**	0.3358**
	All features	0.3621	0.3291	0.4953	0.2008

for an English-speaking profile, a user far from Gaza and Syria geographically and culturally impacts his way of interacting on Facebook. In addition to the two reactions "Sad" and "Angry", we found that the surprise response "Wow" is also strongly present in relevant posts to these two queries. Therefore, when the search query is in English, it is very likely that the user came from an English-speaking community linguistically and culturally, and therefore when we promoted posts carrying the "Wow" reaction at the top of the list of search results, the user was more satisfied with the results.

Our conclusion with respect to this work is twofold:

- The language is a strong cultural element that impacts the user profile on his way of interacting, his interest center and his way of seeing things and his proximity to events and trends of different subjects of life daily (sport, politic, etc).
- Social signals are fruitful for Facebook's information retrieval system. However, they must be exploited by taking into account the user's language and interactional culture in order to better understand how to interpret his interactions in search ranking task. Some users are surprised by the existence of a cruel war in Syria so they can let the reaction "Wow" and a comment expressing his support, his revolt against this evil or his anger and others are aware of this war so they react with an "Sad" or "Angry" reactions.

6 CONCLUSION

To the best of our knowledge, we provide the first comprehensive investigation for the impact of the social features on Arabic Facebook search. This paper proposes a supervised approach of Arabic and English Facebook search based on social features specific to Facebook. Some features are a simple count (Like, Sad, Haha, etc.), while others represent a polarity of comments (positive or negative). We used features selection techniques combined with learning to rank algorithms. The evaluation conducted on the Facebook dataset shows that Random Forests taking as input the features selected by RLF and IGA is the most successful configuration to estimate the relevance ranking of the Arabic and English results, respectively. In addition, LTR algorithms based on the most relevant features according to the selection algorithms are generally better compared to those obtained when the selection algorithms are ignored. Finally, we note that we are aware that the assessment of our approach is still limited. The main weakness of our approach is its dependence on the quality of the sentiment analysis model. An essential treatment step for an effective Arabic SIR is to use a *stemmer* for dialectal Arabic. Further large-scale experiments on other types of datasets are also envisaged. Even with these simple elements, the first results obtained encourage us to invest more in this track. As perspectives of this work in other context, we plan to adapt our approach to other types of information needs such as seeking controversial and contradictory information around specific topics, using pre-processing approaches on the detection of controversies and contradictions [10, 11].

REFERENCES

[1] Nawaf A Abdulla, Nizar A Ahmed, Mohammed A Shehab, and Mahmoud Al-Ayyoub. 2013. Arabic sentiment analysis: Lexicon-based and corpus-based. In *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*. 1–6.

[2] Nawaf A. Abdulla, Nizar A. Ahmed, Mohammed A. Shehab, Mahmoud Al-Ayyoub, Mohammed Naji Al-Kabi, and Saleh Y. Al-Rifai. 2014. Towards Improving the Lexicon-Based Approach for Arabic Sentiment Analysis. *IJITWE* 9, 3 (2014), 55–71.

[3] Nawaf A. Abdulla, Mahmoud Al-Ayyoub, and Mohammed Naji Al-Kabi. 2014. An extended analytical study of Arabic sentiments. *IJBDI* 1, 1/2 (2014), 103–113.

[4] Mohammed N Al-Kabi, Amal H Gigieh, Izzat M Alsmadi, Heider A Wahsheh, and Mohamad M Haidar. 2014. Opinion mining and analysis for Arabic language. *IJACSA) International Journal of Advanced Computer Science and Applications* 5, 5 (2014), 181–195.

[5] Mohamed A. Aly and Amir F. Atiya. 2013. LABR: A Large Scale Arabic Book Reviews Dataset. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 2: Short Papers*. 494–498.

[6] Ismail Badache and Mohand Boughanem. 2014. Harnessing Social Signals to Enhance a Search. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, Poland, August 11-14, 2014 - Volume II*. 303–309.

[7] Ismail Badache and Mohand Boughanem. 2015. A Priori Relevance Based On Quality and Diversity Of Social Signals. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015*. 731–734.

[8] Ismail Badache and Mohand Boughanem. 2017. Emotional Social Signals for Search Ranking. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*. 1053–1056.

[9] Ismail Badache and Mohand Boughanem. 2017. Fresh and Diverse Social Signals: Any Impacts on Search?. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, CHIIR 2017, Oslo, Norway, March 7-11, 2017*. 155–164.

[10] Ismail Badache, Sébastien Fournier, and Adrian-Gabriel Chifu. 2017. Finding and Quantifying Temporal-Aware Contradiction in Reviews. In *Information Retrieval Technology - 13th Asia Information Retrieval Societies Conference, AIRS 2017, Jeju Island, South Korea, November 22-24, 2017, Proceedings*. 167–180.

[11] Ismail Badache, Sébastien Fournier, and Adrian-Gabriel Chifu. 2018. Predicting Contradiction Intensity: Low, Strong or Very Strong?. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*. 1125–1128.

[12] Amira Barhoumi, Nathalie Camelin, and Yannick Estève. 2018. Des représentations continues de mots pour l'analyse d'opinions en arabe: une étude qualitative. In *Traitement Automatique des Langues Naturelles, TALN 2014, 14-18 Mai, 2014, Marseille, France*. 1–9.

[13] Amira Barhoumi, Yannick Estève, Chafik Aloulou, and Lamia Hadrich Belguith. 2017. Document Embeddings for Arabic Sentiment Analysis. In *Proceedings of the First Conference on Language Processing and Knowledge Management, LPKM 2017, Kerkennah (Sfax), Tunisia, September 8-10, 2017*.

[14] Amine Bayoudhi, Hatem Ghorbel, and Lamia Hadrich Belguith. 2015. Sentiment Classification of Arabic Documents: Experiments with multi-type features and ensemble algorithms. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, PACLIC 29, Shanghai, China, October 30 - November 1, 2015*.

[15] Mohamed Reda Bouadjenek, Hakim Hacid, and Mokrane Bouzeghoub. 2016. Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms. *Inf. Syst.* 56 (2016), 1–18.

[16] Sergey Brin and Lawrence Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks* 30, 1-7 (1998), 107–117.

[17] Marco Buijs and Marco R. Spruit. 2014. The Social Score - Determining the Relative Importance of Webpages Based on Online Social Signals. In *KDIR 2014 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, Rome, Italy, 21 - 24 October, 2014*. 71–77.

[18] Beate Navarro Bullock, Andreas Hotho, and Gerd Stumme. 2018. Accessing Information with Tags: Search and Ranking. In *Social Information Access - Systems and Technologies*. 310–343.

[19] Sergiu Chelaru, Claudia Orellana-Rodriguez, and Ismail Sengör Altingövde. 2014. How useful is social feedback for learning to rank YouTube videos? *World Wide Web* 17, 5 (2014), 997–1025.

[20] Abdelghani Dahou, Shengwu Xiong, Junwei Zhou, Mohamed Houcine Haddoud, and Pengfei Duan. 2016. Word Embeddings and Convolutional Neural Network for Arabic Sentiment Classification. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*. 2418–2427.

[21] Firas Damak. 2014. *Étude des facteurs de pertinence dans la recherche de microblogs. (Study of salient factors for microblog search).* Ph.D. Dissertation. Paul Sabatier University, Toulouse, France.

[22] Alaadatasets presented indatasets presented indatasets presented indatasets presented indatasets presented in El-Halees. 2011. Arabic opinion mining using combined classification approach. (2011).

[23] Hady ElSahar and Samhaa R. El-Beltagy. 2015. Building Large Arabic Multi-domain Resources for Sentiment Analysis. In *Computational Linguistics and Intelligent Text Processing - 16th International Conference, CICLing 2015, Cairo, Egypt, April 14-20, 2015, Proceedings, Part II*. 23–34.

[24] Noura Farra, Elie Challita, Rawad Abou Assi, and Hazem M. Hajj. 2010. Sentence-Level and Document-Level Sentiment Mining for Arabic Texts. In *ICDMW 2010, The 10th IEEE International Conference on Data Mining Workshops, Sydney, Australia, 13 December 2010*. 1114–1119.

[25] Mark A. Hall and Geoffrey Holmes. 2003. Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. *IEEE Trans. on Knowl. and Data Eng.* 15, 6 (2003), 1437–1447.

[26] Hossam S. Ibrahim, Sherif M. Abdou, and Mervat Gheith. 2015. Sentiment Analysis For Modern Standard Arabic And Colloquial. *CoRR* abs/1505.03105 (2015).

[27] Thorsten Joachims. 2006. Training linear SVMs in linear time. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, August 20-23, 2006*. 217–226.

[28] Wessel Kraaij, Thijs Westerveld, and Djoerd Hiemstra. 2002. The Importance of Prior Probabilities for Entry Page Search. In *SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 11-15, 2002, Tampere, Finland*. 27–34.

[29] Zhunchen Luo, Miles Osborne, Sasa Petrovic, and Ting Wang. 2012. Improving Twitter Retrieval by Exploiting Structural Information. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada*.

[30] Craig Macdonald and Iadh Ounis. 2006. Voting for candidates: adapting data fusion techniques for an expert search task. In *Proceedings of the 2006 ACM CIKM International Conference on Information and Knowledge Management, Arlington, Virginia, USA, November 6-11, 2006*. 387–396.

[31] Matteo Magnani, Danilo Montesi, and Luca Rossi. 2012. Conversation retrieval for microblogging sites. *Inf. Retr.* 15, 3-4 (2012), 354–372.

[32] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*.

[33] Saif Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Proceedings of the 7th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2013, Atlanta, Georgia, USA, June 14-15, 2013*. 321–327.

[34] Ananth Mohan, Zheng Chen, and Kilian Q. Weinberger. 2011. Web-Search Ranking with Initialized Gradient Boosted Regression Trees. In *Proceedings of the Yahoo! Learning to Rank Challenge, held at ICML 2010, Haifa, Israel, June 25, 2010*. 77–89.

[35] Mahmoud Nabil, Mohamed A. Aly, and Amir F. Atiya. 2015. ASTD: Arabic Sentiment Tweets Dataset. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*. 2515–2519.

[36] Valeria Orso, Tuukka Ruotsalo, Jukka Leino, Luciano Gamberini, and Giulio Jacucci. 2017. Overlaying social information: The effects on users' search and information-selection behavior. *Inf. Process. Manage.* (2017).

[37] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP 2002, Philadelphia, PA, USA, July 6-7, 2002*.

[38] Alec Radford, Rafal Józefowicz, and Ilya Sutskever. 2017. Learning to Generate Reviews and Discovering Sentiment. *CoRR* abs/1704.01444 (2017).

[39] Eshrag Refaee and Verena Rieser. 2014. An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014*. 2268–2273.

[40] Eshrag Refaee and Verena Rieser. 2016. iLab-Edinburgh at SemEval-2016 Task 7: A Hybrid Approach for Determining Sentiment Intensity of Arabic Twitter Phrases. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*. 474–480.

[41] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*. 1631–1642.

[42] Peter D. Turney. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA*. 417–424.

[43] Thijs Westerveld, Wessel Kraaij, and Djoerd Hiemstra. 2001. Retrieving Web Pages Using Content, Links, URLs and Anchors. In *Proceedings of The Tenth Text REtrieval Conference, TREC 2001, Gaithersburg, Maryland, USA, November 13-16, 2001*.

[44] Zhen Yue and Daqing He. 2018. Collaborative Information Search. In *Social Information Access - Systems and Technologies*. 108–141.