What Sentiments Can Be Found in Medical Forums?

Marina Sokolova
CHEOR Research Institute
University of Ottawa
sokolova@uottawa.ca

Victoria Bobicev
Technical University of Moldova
vika@rol.md

Abstract

In this work we present sentiment analysis of messages posted on a medical forum. We categorize posts, written in English, into five categories: *encouragement, gratitude, confusion*, *facts*, and *facts + sentiments*. Our study applies a manual sentiment annotation, affective lexicons in its sentiment analysis and machine learning classification of sentiments in these texts. We report empirical results obtained from analysis of 752 posts dedicated to infertility treatments. Our best results improve multi-class sentiment classification of online messages (*F-score* = 0.518, *AUC*= 0.685).

1 Introduction

User-friendly Web 2.0 technologies encourage the general public actively participate in the creation of the Web content. Blogs, social networks, message boards reach out to a global community of the Web users. The online texts discuss personal experience and convey sentiments and emotions of the authors. These emotion-rich posts are known to be important in setting interaction patterns among members of online communities as emotion-rich text has a strong influence on a public mood (Allan, 2005). Subjective information posted by a user may affect subjectivity in posts written by other users (Zafarani et al 2010).

Studies of online sentiments and opinions can help in understanding of sentiments and opinions of the public at large. Such understanding is especially important for the development of public policies whose success greatly depends on public attitudes. Among major policy issues (e.g., education, internal and foreign affairs), health care policies are those that directly affect everybody and cause many online discussions. A 2011 survey of the US population estimated that 59% of all adults have looked online for information about health topics such as a specific disease or treatment (Fox 2011). Reproductive technologies belong to a group of hotly debated health care issues in the modern societies (Zillen 2011). The systematic review of 19 studies from 1999-2009 listed several reasons for the use of medical forums: a) information searching - to learn about psychological, physical and social aspects of available treatments, evaluations of alternative treatments; b) in seeking emotional support - anonymous communication, immediate and constant community access, easy contact to peers.

We analyzed sentiments expressed by participants of *In Vitro Fertilization* (*IVF)* medical forum.[1] This forum brings together women who use IVF treatments with the hope to conceive. For the empirical analysis, we selected 752 posts that covered 74 topics related to IVF (e.g., *Over 40 and pregnant or trying to be, Odds of getting pregnant naturally on a cancelled IVf cycle, Going for a second opinion*). Starting with several possible sentiments, we finally categorized text into *encouragement, gratitude, confusion*, *facts + encouragement,* and *facts*. Texts in which the annotators disagreed on a class label were labeled as *uncertain*.

In the analysis, we applied a three-fold approach. First, we manually annotated the messages and then analyzed agreement between annotators. Second, we used affective lexicons for the sentiment analysis of the data. Next, we identified a multi-class classification problem and ran experiments to automatically classify posts into the five categories. The obtained results show a high agreement between the annotators (*Fleiss Kappa* = 0.73) and significant accuracy improvement over baseline (*F-score* = 0.518, *AUC*= 0.685 vs. the baseline *F-score* = 0.118, *AUC*= 0.491).

[1] http://ivf.ca/forums

Proceedings of Recent Advances in Natural Language Processing, pages 633–639,
Hissar, Bulgaria, 7-13 September 2013.

2 Related works

Sentiment analysis has become a major research field in Text Data Mining and Computational Linguistics. Machine Learning (ML) methods, affective lexicons, and Natural Language Processing (NLP) apparatus are used to classify text units (e.g., words, sentences, paragraphs) into sentiment categories (Taboada et al, 2011). Availability of on-line data prompted sentiment analysis of user-written messages posted on the Web (Dodds et al. 2011; Thelwall at al., 2010; Jansen et al. 2009; Chmiel et al 2011). In this study, we worked with online messages posted on a medical forum. Hence a message is the main text unit on the Web forums we decided to keep it as our text unit.

Although empirical evidence strongly supports the importance of emotions in health-related messages (Pennebaker and Chung, 2006), there are few studies of the relationship between a subjective language and online discussions of personal health (Smith 2011). 16 categories of opinions and emotions in tweets were presented in (Chew and Eysenbach, 2010). The extraction method looked for tweets with references to H1N1 and its synonyms. However, numerical evaluation of the method was not reported by the authors. Sokolova and Bobicev (2011) studied positive and negative opinions and positive and negative sentiments in the health-related sci.med messages from *20 NewsGroups*.[2] For sentiments, Support Vector Machines obtained the best *Fscore* (70.8%). Sentiments in short health-related messages were studied in (Bobicev et al, 2012). The authors analyzed positive, negative and neutral sentiments expressed in tweets that discuss personal health. The Twitter data, however, contained a limited number of health-related tweets: among 409 analyzed tweets, only 124 tweets discussed personal health. In the current work, we obtained the results on 752 health-related messages, hence, gathered stronger empirical evidence.

Sentiment research often uses lexicons where words are assigned with opinion, sentiment, and emotion categories (Wilson et al, 2005; Strapparava et al, 2006; Strapparava and Mihalcea, 2008). The most popular resources are SentiWordNet[3], WordNetAffect[4] and the Subjectivity lexicon[5]. Although there was a study on the use of affective lexicons in discussion of prescriptive drugs (Goeuriot et al, 2012), to the best of our knowledge, there were no previous applications of affective lexicons to sentiment analysis of online discussions of personal health. In the current work, we experimented with the application of four affective lexicons in the sentiment analysis of online discussions of personal health.

Few publications focused on manual sentiment annotation of online messages. Topic-specific opinions in blogs were evaluated in Osman et al., (2010). Agreement among seven manual annotators was computed for five classification categories, including positive, negative, mixed opinions and non-opinionated and non-relevant categories. Sokolova and Bobicev (2011) evaluated concordance of the manual annotation of messages posted on a medical forum. The results show that annotators more strongly agree on what sentences do *not* belong to positive or negative subjective categories then on what sentences *do* belong to those categories. Bobicev et al (2012) used multiple annotators to categorize tweets into positive and negative sentiments and neutral tweets. The authors found that in annotation of health-related tweets annotators more strongly agreed on negative sentiments than on positive ones (p_{pos}= 0.22, p_{neg} = 0.35). The opposite was true for tweets that did not discuss personal health: annotators more strongly agreed on positive sentiments than on negative ones. Our current study addresses manual assignment of health-related texts with several classification labels.

3 Data

Our current research focuses on sentiment identification in messages posted on IVF forums. Such forums belong to an infertility outreach resource community created by prospective, existing and past IVF (In Vitro Fertilization) patients. The IVF.ca website includes forums: *Cycle Friends*, *Expert Panel*, *Trying to Conceive*, *Socialize*, *In Our Hearts*, *Pregnancy*, *Parenting*, and *Administration*.[6] Every forum hosts a few sub-forums, e.g. the *Cycle Friends* forum has six sub-forums:

[2] http://qwone.com/~jason/20Newsgroups/

[3] http://sentiwordnet.isti.cnr.it/

[4] http://wndomains.fbk.eu/wnaffect.html

[5] http://mpqa.cs.pitt.edu/#subj_lexicon

[6] www.ivf.ca/forums

Introductions, *IVF/FET/IUI Cycle Buddies*, *IVF Ages 35+*, *Waiting Lounge*, *Donor & Surrogacy Buddies*, and *Adoption Buddies*. On every subforum, topics are initiated by the forum participants. Depending on the interest among participants, a different number of messages is associated with each topic, e.g., *Human growth hormone & what to expect* has 120 messages posted from Oct 2012, while *Over 40 and pregnant or trying to be* has 3,455 messages posted from May 2010.

We wanted the forum to represent many discussions, and so forums were selected to ensure a high number of topics and large number of posts. The *IVF Ages 35+* sub-forum[7] satisfied both requirements.

In July 2012, it had 510 topics and 16388 messages. At this point, we discharged the largest four topics containing 7498, 2823, 1131 and 222 posts respectively; we will indentify the shortest topics and discharge them later on. Figure 1 presents the statistics for the rest of the topics in this subforum, i.e. the largest four topics are not shown in the chart. Topics are sorted by the number of posts in them in descending order. The topic's rank is its number in the sorted list.

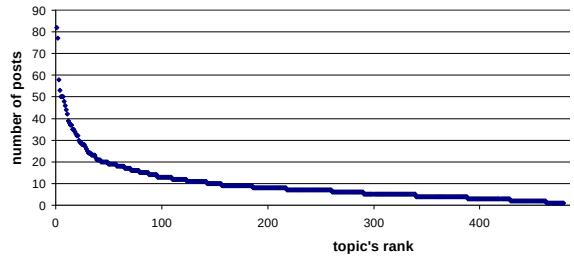

Figure 1: Number of posts per topic in the *IVF Ages 35+* sub-forum

Among the remaining 506 topics, we looked for those where the forum participants discussed only one theme. A preliminary analysis showed that discussions with ≤ 20 posts satisfied this condition. Also, we wanted discussions be long enough to form a meaningful discourse. This condition was satisfied when discussion had ≥ 10 messages. As a result, for further analysis, we analyzed 74 topics with 10 - 20 posts, with an average 12.5 messages per topic. Most of the topics had a similar structure:

a) a participant started the theme with a post;

b) the initial post usually contained some information about the participant's problem, expressed worry, concern, uncertainty and a request for help to the other forum participants;

c) the following posts:

i) provided the requested information by describing their similar stories, knowledge about treatment procedures, drugs, doctors and clinics, or

ii) supplied moral support through compassion, encouragement, wishing all the best, good luck, etc.

d) the participant who started the topic often thanked other contributors and expressed appreciation for their help and support.

4 Manual Annotation

4.1 Model

Annotation of subjectivity can be centered either on the perception of a reader (Strapparava, Mihalceal, 2008) or the author of a text (Balahur, Steinberger, 2009). In the current work, we aimed to detect sentiments conveyed by posts of the forum participants. Hence, we opted for the reader perception model and asked annotators to analyze the topic's sentiment as it was addressed toward the other forum participants.

We asked annotators to label the post with the dominant sentiment. Posts that combined factual information and sentiments usually expressed encouragement for specific participants, hence we suggested the label "*facts +encouragement"* for that category.

4.2 Identification of sentiments.

We wanted to know what types of sentiments were dominant in these forums and how these sentiments influence each other. Previously, analysis of the topics' content revealed that most posts referred to sharing personal experiences, provision of information or advice, expressions of gratitude/friendship, chat, requests for information, and expressions of universality (e.g. "*we're all in this together"*) (Malik, Coulson, 2010). Hypothesizing that binary sentiment categories (e.g., positive and negative polarity), would be too general and could not adequately cover emotions expressed in health-related messages, we intended to build a set of sentiments that

[7] http://ivf.ca/forums/forum/166-ivf-ages-35/

1. contains sentiment categories specific for posts from medical forums, and
2. makes feasible the use of machine learning methods for automate sentient detection.

To identify such a set, we asked annotators to read several topic discussions and describe sentiments expressed by the forum participants and the sentiment propagation within these discussions. More specifically, the annotators were told to indicate sentiments in sequences. For example, we asked annotators to answer groups of questions:

- What sentiment was expressed in the first post in the topic? How were the sentiments of the following posts affected by the initial sentiment?
- How long did an expressed sentiment last in the topic? If it was replaced by another one, how did the replacement happen?
- Did the participants joining the discussion try to change the previous sentiments? Did the participants succeed in such attempts?

We asked annotators not to mark descriptions of symptoms and diseases as subjective; in many cases they appear in the post as objective information for other forum participants that have encountered similar issues. In such cases only the author's sentiments toward other participant should be taken into consideration. For example, `I have had a few days now with heartburn/reflux - could be stress, a little achy tummy/pelvic and a tired aching back. More waiting, but getting more hopeful` is a description of symptoms and should not be annotated as subjective. In contrast, `I hope your visit with us infertilies is short and sweet and you get that baby soon!!!` exposes the author's sentiment towards another person.[8]

The data annotation was carried on by the Master's students as their practical work for the course "Semantic Interpretation of Text". The students already completed courses on "Computational Linguistics" and "Natural Language Processing". Based on the quality of annotations, eight annotators were selected after the first phase of the sentiment analysis. Most annotators already had experience in text annotation. Each annotator independently annotated a set of topics. Each annotator filled in a short questionnaire for every analyzed topic. After that, we merged and summarized all questionnaires.

4.3 The annotation scheme

Based on the responses to the questionnaires, we built three groups of sentiments:

1. **confusion**, which included worry, concern, doubt, impatience, uncertainty, sadness, angriness, embarrassment, hopelessness, dissatisfaction, and dislike;
2. **encouragement**, which included cheering, support, hope, happiness, enthusiasm, excitement, optimism;
3. **gratitude**, which included thankfulness.

A special case was presented by expressions of *compassion*, *sorrow*, and *pity* which did not appear individually but appeared in conjunction with encouragement; we treated them as a part of encouragement.

Also, we identified two types of posts with factual information: *facts* and *facts + encouragement*. Posts were marked as *facts* if they delivered factual information only. Posts were marked as *facts + encouragement* when they contained factual information supplemented by short emotional expressions; those expressions almost always conveyed encouragement ("*hope, this helps*", "*I wish you all the best*", "*good luck*").

As a result, our annotation schema was implemented as follows:

(a) annotation was performed on a level of individual posts; annotators were asked to select the most dominant sentiment in the whole post; descriptions of symptoms or diseases were omitted from the sentiment annotation;

(b) every post was marked with only one label; at this stage we did not aim to identify interrelations between sentiments; this task is delegated to the next stage of our study;

(d) finally, every post was labeled by two annotators.

We evaluated agreement between the annotators by using Fleiss Kappa (Nichols et al, 2010), a measure that evaluates agreement for a multiclass manual labeling.

$$Fleiss\ Kappa = (P - P_{class})/(1-P_{class})$$

where P is an average agreement per a post and P_{class} is an average agreement per a class. For a five-class problem, the annotators achieved a high agreement: Fleiss Kappa = 0.73 which indicates a strong agreement (Osman et al, 2010).

Preparing our data for the machine learning experiments we assigned the five category labels

[8] All examples preserve original spelling and grammar.

only to posts that both annotators labeled with the same label, e.g., if a post was labeled *encouragement* by two annotators it was put into the *encouragement* category. We introduced a new class *uncertain* for the posts labeled with two different labels. The final number of posts per class was:
Encouragement – 206, *Gratitude* – 88, *Confusion* – 48, *Facts* – 187, *Facts + Encouragement* - 73, and *Uncertain*– 150; total – 752 posts.

5 HealthAffect

To the best of our knowledge, WordNet-Affect[9] is the only affective lexicon with a highly detailed hierarchy of sentiments (Strapparava et al 2006). Other affective lexicons assign words with positive and negative polarity labels only (e.g., SentiWordNet (Baccianella et al. 2010), Bing Liu's Opinion Lexicon [10] (Liu, 2010), MPQA subjectivity lexicon (Wiebe et al., 2005)).

However, comparison of the post vocabulary with WordNet-Affect words revealed that very few words from WordNet-Affect appeared in any given post's text. Consider a dialogue from Example 1.

Example 1. **post_id_140772** The test is Positive!!! I'm giving you dancing banana's.
post_id_140789 I'm thinking that 64 sounds positive to me! I second Hopeful Flyer with the dancing bananas and raise her a for a BFP.
post_id_141266 thanks for your wishes The nurse at Edmonton called me and wants me to re-test
post_id_141340 yay! congrats! best of luck on test!
post_id_141455 Baby dust to you. Fingers crossed. Keep Positive.

In Example 1, there was only one word - *positive* - which was found in WordNet-Affect; thanks, congrats!, best of luck, Fingers crossed were not found in the WordNet-Affect dictionary. On the other hand, some WordNet-Affect words were used in posts in the senses not related to sentiments (e.g. *get*, *move*, *close*, *cold*).

As those matching result were unsatisfactory, we created a specific lexicon which we named HealthAffect. To build HealthAffect, we adapted the Pointwise Mutual Information (PMI) of *word*1 and *word*2 (Turney, 2002):

$$PMI(word1, word2) = \log_2(p(word1 \& word2)/(p(word1)\,p(word2)))$$

First, we created a list of all words, bigrams and trigrams of words with frequency ≥ 5 from the unambiguously annotated posts (i.e., we omitted posts marked as *uncertain*). This was a list of candidates (aka *phrases*) to be included in our HealthAffect lexicon. Note that the Part-of-Speech tagging would be ineffective due to a high volume of textual noise (e.g., incomplete sentences, InternetSpeak jargon, loose grammar).

Next, for each class, we calculated PMI(*phrase*, *class*) as

$$PMI(phrase, class) = \log_2(p(phrase \text{ in } class)/(p(phrase)\,p(class))).$$

Finally, we calculated Semantic Orientation (SO) for each phrase and for each class as

$$SO(phrase, class) = PMI(phrase, class) - \Sigma\, PMI(phrase, other\_classes)$$

where *other_classes* are all the classes except for the class that Semantic Orientation is calculated for.

After all the possible SOs were computed, each HealthAffect candidate was assigned with the class that corresponded to its maximum SO. Consequently, each candidate was considered an indicator of the class that provided it with the maximum SO. It should be noted that each class got different numbers of indicative candidates. From 459 trigrams with frequency ≥ 5, 46 had their maximum SO for *encouragement*, 40 - for *gratitude*, 139 - for *confusion*, 95 - for *facts* and 139 for *facts + encouragement*.

For each class, we sorted all potential N-grams in decreasing order of SO and selected the equal number of N-grams to represent each class in the lexicon. The number of N-grams was determined as ½ of the minimum *per class* number of N-grams; for example, we used only 20 (=40:2) top trigram indicators for each class. Similarly, we selected 50 bigrams and 25 unigrams and added them to the lexicon.

A direct matching of HealthAffect to unambiguously annotated posts gave the following results:

- lexicon annotation matched the human annotation – 420 posts;
- lexicon annotation did not match the human annotation – 182 posts.

Thus, lexicon-based annotation matched 70% of unambiguously annotated posts. Therefore we used the created lexicon in Machine Learning experiments.

[9] http://wndomains.fbk.eu/wnaffect.html

[10] www.cs.uic.edu/~liub/FBS/opinion_lexicon_English.rar

6 Machine Learning Experiments

We used personal pronouns, short words, the WordNetAffect terms and the HealthAffect lexicon in four data representations:

- all semantic features (AllSem),
- WordNetAffect and pronouns features (WNAP),
- WordNetAffect features (WNA).
- HealthAffect lexicon (HAL)

We used Naïve Bayes (NB) and K-nearest neighbor (KNN) to classify the messages into 6 classes.

We assessed the learning methods by computing multi-class *Precision (Pr), Recall (R), F-score (F)* and *Accuracy Under the Curve (AUC).* We used 10-fold cross-validation to select the best classifier. Labeling all examples as the majority class gave the baseline for the performance evaluation: *Pr*= 0.075, *R* = 0.274, *F* = 0.118, *AUC* = 0.491. Table 1 and Table 2 report the empirical results.

NB results				
Features	*Pr*	*R*	*F*	*AUC*
AllSem	0.408	0.427	0.397	0.685
WNAP	0.324	0.395	0.333	0.661
WNA	0.322	0.350	0.303	0.605
HAL	0.527	0.541	0.518	0.799

Table 1: NB results in 6-class classification.

KNN results				
Features	*Pr*	*R*	*F*	*AUC*
AllSem	0.330	0.342	0.310	0.598
WNAP	0.287	0.319	0.284	0.591
WNA	0.279	0.322	0.275	0.571
HAL	0.377	0.376	0.340	0.619

Table 2: KNN results in 6-class classification.

Empirical evidence shows that while solving the multi-class classification problem, we significantly improved over the baseline ($P < 0.01$, paired t-test). HealthAffect provided a more accurate classification of sentiments, and NB outperformed KNN on all the data representations. However, for NB, the difference between the best and the worst F-score was as high as 60%, whereas for KNN the difference was < 10%.

7 Conclusions and Future Work

In this work, we have presented the sentiment analysis of messages posted on medical forums. We stated the sentiment analysis as a multi-class classification problem in which posts were classified into *encouragement, gratitude, confusion, facts, facts + encouragement* and *uncertain* categories. We applied the reader-centered manual annotation and achieved a strong agreement between the annotators: *Fleiss Kappa* = 0.73.

Sentiment analysis of online medical discussions differs considerably from the traditional studies of sentiments in consumer-written product reviews, financial blogs and political discussions opinion detection. While in many cases positive and negative sentiment categories are enough, such dichotomies are not sufficient for medical forums. The same can be said about the existing sentiment and affective lexicons: their general terms and labels do not adequately serve for the analysis of medical posts. Thus, new lexical resources sensitive to this specific domain should be created. We presented an ad-hoc method of the lexicon creation which is comparatively easy to implement. We have shown that the lexicon, which we call HealthAffect, provided the best accuracy in machine learning experiments. However, as many other lexical resources, the lexicon requires manual review and filtering. In the future, we plan to analyze and optimize this lexicon manually.

We used two algorithms, NB and KNN, to solve a multi-class sentiment classification problem. The probability-based NB demonstrated a better performance than KNN. The best F-score was achieved when posts were represented through HealthAffect, an affective lexicon built to identify sentiments in health-related online posts.

We present this work as the first phase of our analysis of medical forums. Our long term goal is to analyze health-related online discourses. We are interested in sentiment interaction, flow and propagation in these dialogues. To achieve this goal, we need a reliable tool for sentiment detection specifically in heath-related online texts.
In the future, we aim to annotate more texts, enhance and refine our lexicon and achieve reliable automated sentiment detection in health-related messages. We plan to use the results obtained in this study to perform analyses of health-related discussions on medical forums related to highly debatable health care policies.

Acknowledgements

This work was in part supported by NSERC Discovery grant. The authors thank Brian Dewar for his assistance with the manuscript editing.

References

Allan, K. 2005. *Explorations in Classical Sociological Theory*: Seeing the Social World. Pine Forge Press, 2005.

Baccianella, S., A. Esuli, and F. Sebastiani. 2010. *SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining.* Proceedings of the 7th LREC, 2200-2204.

Balahur, A. and R. Steinberger. 2009. *Rethinking Sentiment Analysis in the News: from Theory to Practice and back.* Proceedings of the 1st Workshop on Opionion Mining and Sentiment Analysis, 2009.

Bobicev, V., M, Sokolova, Y. Jaffer, D. Schramm. 2012. *Learning Sentiments from Tweets with Personal Health Information*. Proceedings of Canadian AI 2012, p.p. 37–48, Springer, 2012.

Chen, W. 2008. *Dimensions of Subjectivity in Natural Language (Short Paper)*. In Proceedings of ACL-HLT, 2008.

Chew, C. and G. Eysenbach. 2010. *Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak*. PLoS One, 5(11), 2010.

Chmiel, A., J. Sienkiewicz, M. Thelwall, G. Paltoglou, K. Buckley, A. Kappas, J. Holyst. 2011. *Collective Emotions Online and Their Influence on Community Life*. PLoS one, 2011.

Dodds, P., K. Harris, I. Kloumann, C. Bliss, C. Danforth. 2011. *Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter*. PLoS ONE, 6, e26752, 2011.

Fox, S. 2011. *The Social Life of Health Information*. Pew Research Center's Internet & American Life Project,http://pewinternet.org/Reports/2011/Social-Life-of-Health-Info.aspx

Goeuriot, L., J. Na, W. Kyaing, C. Khoo,Y. Chang, Y. Theng and J. Kim. 2012. *Sentiment lexicons for health-related opinion mining*. Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, p.p. 219 – 225, ACM.

Jansen, B.J., Zhang, M., Sobel, K., and Chowdury, A. 2009. *Twitter power: Tweets as electronic word of mouth.* Journal of the American Society for Information Science and Technology, 60(11), 2169-2188, 2009.

Liu B. 2010. *Sentiment Analysis and Subjectivity*. Handbook of Natural Language Processing, Second Edition, 2010.

Malik S. and N. Coulson. 2010. *Coping with infertility online: an examination of self-help mechanisms in an online infertility support group*. Patient Educ Couns, vol. 81, no. 2, pp. 315–318, Nov. 2010.

Nichols, T., P. Wisner, G. Cripe, and L. Gulabchand. 2010. *Putting the Kappa Statistic to Use.* Qual Assur Journal, 13, p.p. 57-61, 2010.

Oakes, M. 2005. *Statistics for Corpus Linguistics*. Edinburgh University Press.

Osman, D., J. Yearwood, P. Vamplew. 2010. *Automated opinion detection: Implications of the level of agreement between human raters.* Information Processing and Management, 46, 331-342, 2010.

Pennebaker, J. and Chung, C. 2006. *Expressive Writing, Emotional Upheavals, and Health*. Handbook of Health Psychology, Oxford University Press.

Smith, C. 2011. *Consumer language, patient language, and thesauri: A review of the literature*. Journal of the Medical Library Association, 99(2), 135– 144, 2011.

Sokolova, M. and V. Bobicev. 2011. *Sentiments and Opinions in Health-related Web Messages.* Recent Advances in Natural Language Processing, p.p. 132- 139, 2011.

Strapparava, C. and R. Mihalcea. 2008. *Semeval-2007 task 14: Affective text.* Proceedings of the 2008 ACM symposium on Applied computing, 2008.

Strapparava, C., A. Valitutti, and O. Stock. 2006. *The affective weight of the lexicon.* Proceedings of the 5th International Conference on Language Resources and Evaluation, pages 474-481, 2006.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll and M. Stede. 2011. *Lexicon-Based Methods for Sentiment Analysis*. Computational Linguistics 37 (2): 267-307.

Thelwall, M., K. Buckley, and G. Paltoglou. 2010. *Sentiment in Twitter events.* Journal of the American Society for Information Science and Technology, 62(2), 406-418, 2010.

Turney, P.D. 2002. *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*. Proceedings of ACL'02, Philadelphia, Pennsylvania, pp. 417-424.

Wiebe, J., T. Wilson, and C. Cardie. 2005. *Annotating expressions of opinions and emotions in language.* Language Resources and Evaluation 39: 165-210.

Wilson, T., J. Wiebe, and P. Hoffmann. 2005. *Recognizing contextual polarity in phrase-level sentiment analysis.* Proceedings of EMNLP 2005. Association for Computational Linguistics

Zafarani, R., W. Cole, and H. Liu. 2010. *Sentiment Propagation in Social Networks: A Case Study in LiveJournal*. Advances in Social Computing (SBP 2010), pp. 413–420, Springer

Zillen, N. 2011. Internet Use of Fertility Patients: A Systemic Review of the Literature. *Journal of Reproductive Medicine and Endocrinology*, 8(4): 281–287.