Professor Maria Teresa PAZIENZA, PhD
University of Rome Tor Vergata
Italy
E-mail: pazianza@info.uniroma2.it
Professor Ion LUNGU, PhD
E-mail: ion.lungu@ie.ase.ro
Alexandra TUDORACHE , PhD Candidate
E-mail: tudorache@info.uniroma2.it
The Bucharest Academy of Economic Studies

FLAMES RECOGNITION FOR OPINION MINING

***Abstract**. The emerging world-wide e-society creates new ways of interaction between people with different cultures and backgrounds. Communication systems as forums, blogs, and comments are easily accessible to end users. In this context, user generated content management revealed to be a difficult but necessary task. Studying and interpreting user generated data/text available on the Internet is a complex and time consuming task for any human analyst.*

This study proposes an interdisciplinary approach to modelling the flaming phenomena (hot, aggressive discussions) in online Italian forums. The model is based on the analysis of psycho/cognitive/linguistic interaction modalities among web communities' participants, state-of-the art machine learning techniques and natural language processing technology. Virtual communities' administrators, moderators and users could benefit directly from this research. A further positive outcome of this research is the opportunity to better understand and model the dynamics of web forums as the base for developing opinion mining applications focused on commercial applications.

***Keywords:** flaming, flame wars, web forums, opinion mining, natural language processing, machine learning.*

JEL Classification: C10, C31, C38, D80, D83, L86

Introduction

Nowadays, most human communication takes place online: forums, blogs, and comment systems. User content is generated at a really high pace and in large volumes. A unique psycho/cognitive/linguistic model of these interactions is still missing, while the demand to process such data for several different applications emerged. Humans have a variety of channels of communication: verbal (face to face), non-verbal (paralanguage, body language, touch, glance), written (novels, letters, emails, posts) and media channel (images, videos, sounds). [1]

In Internet most communication misses the non verbal part and this generates/provoke information loss. Pictures, videos, and music are used seldom in forums, mailing lists and comment systems. In blogs and other less dynamic environments sometimes different media are used to underline or illustrate ideas. One example of miscommunication generated by the transmission media are the flame wars in forums.

Our study focuses on modelling the complex flaming phenomena (hot, aggressive discussions) in Italian forums.

Forum dynamics are complex and depend on different elements. Often in forums there are users from every corner of the world with different degrees of knowledge of the forum language and with different cultures and backgrounds, ages and gender. Furthermore, user behaviour is determined by several factors that often are not observable (personal problems, mood), while we can analyse only the visible results: the written posts. Mabry [14] showed that one expression written on a friendly topic can have a colloquial meaning but on a risky topic can promote flames, while early mediation has a positive effect in online communication.

In such dynamic environment it is better to prevent flames than to be forced to close discussions and even ban (exclude) users. Therefore, in our analysis we considered three types of discussions: flames, no flames and risky topics. Risky topics are situated between flames and no-flames and have features common to both flames and normal discussions.

We identified several specific features and hypothesized that both the inner structure of board language and user behaviour could be good discriminators between flames, risky topics and normal discussions. Each forum is a distinct community with its own users, staff (moderators, administrators) and rules. This closed world model impacts greatly on language, behaviour and consequently on flames characteristics: there is not a universal model of flaming!

Therefore the analysis should be conducted on data extracted from the same board or same board or section. The data used for our research was collected from the politics section of a real life Italian forum (its name will not be published for privacy reasons).

The study could help moderators, forum administration and users to participate to a more friendly, flames free community open to all ages, preventing legal issues correlated to offensive language, privacy and minor abuse.

More applications could benefit from this research as**:** identify trolls, posts authors on other forums, detect hot topics, email filtering, and identify threats: angry, subversive persons or groups of persons in political or corporate forums and social networks.

Furthermore this study has a great economic value for the study of different products and brands impact in online communities (hot topic/brand/feature identification). The intensity of language and the risk for flaming are proportional to the interest surrounding a product. (E.g. the discussions between the enthusiasts for concurrent brands). Moreover, being able to extract linguistic features represents an additional value to identify opinions.

In the first section we will try to answer to the question "why flames occur" introducing some notions of psychology regarding general communication model, web communication model, group behaviour in online communities, disinhibition effect and personal language.

In the second section it will be analyzed "how flames occur" including the way people express concepts and feelings in forums, user behaviour and their interactions.

In the third section our approach to flames modelling and identification will be widely discussed, including corpus, the proposed model and algorithm.

To conclude, experimental setup and results will be discussed as well as the related research and future work directions. This research also demonstrates that is possible to model and have a good recognition of flames even with a small biased corpus.

Online Communication Model

To better understand the complex phenomena that lie behind flaming in web communities we need to study several aspects of the communication theory and the new media interaction with human groups. Our study explores in a unique context the psychological, linguistic and mathematical aspects of communication. We should address:

a) The general communication model;
b) How it changes in the World Wide Web;
c) The interaction of social groups in computer mediated communication focussing on forums;
d) The online disinhibition effect;
e) The personal language;
f) The flaming phenomena.

General Communication Model

In 1949 Shannon and Weaver [23] proposed the first major communication model at Bell Laboratories. They considered a simple, linear schema for information transmission that includes five elements: an information source, a transmitter, the channel, the receiver and the destination. In human communication this can be illustrated as follows: Idea in A's mind → a formulated message (e.g. sentence) → transfer mechanism (e.g. speech and hearing; written message and reading) → message decoding→ idea in B's mind. Each transformation contributes to the probability of information loss and misunderstanding, which are the main fuel for arguments.

Mehrabian [16] analyzed the ways in which people communicate. There are three major parts in human communication: 55% of message is represented by facial expression; 38% by paralinguistics (pitch, loudness, rate, and fluency); 7% lies in the spoken words. Mehrabian's model shows that body language and tonality is a more accurate indicator of emotions and meaning than the words themselves.

Gudykunst [8] showed that different cultures have different ways of communicating: North-Americans relay more on the meaning of words while Latin Europeans relay more on non-verbal communication and behaviour. It is the so called low or high context communication.

Web Communication Model

In the World Wide Web, the general communication model suffers some changes, mostly determined by the specifics of the transmission medium. Online communication takes place asynchronously in forums, blogs, emails, comment systems and mailing lists.

Communication barriers as: the absence of facial expression and paralinguistic information, different users' background, language knowledge and level of education determine that we receive only a part of the message and even less emotional information. It is up to each person to interpret the message depending of his mood, background, imagination and most important context. Riordan, M. A. and Kreuz, R. J. [22] showed that in this context new specific communication encoding has emerged as: emoticons, abbreviations and a specialized English vocabulary.

Forums as Social Groups

Forums have the dimension of a social group. Peck [20] shows in his research that are four stages of community building: pseudo-community, chaos, emptiness and true community. In the first stage the members act as if they understand each other perfectly. Then, they start to show their disagreements and differences. At the last stage, in a true community, people attain a great level of tacit understanding and empathy. Even heated discussions don't get offensive and bitter, and individual motivation is never questioned.

The dynamics of forums create an unusual situation in which all four stages of community building are present at the same time. There is a stable nucleus of users that are already a community, but at the same time new members arrive and some of the old members leave.

Online Disinhibition Effect

People write and behave in cyberspace in ways that they would not normally use in face-to-face situations. Suler [26] calls this the "disinhibition effect" with two different effects one positive and the one negative. As a positive effect, people tend to open up and to communicate. They share personal emotions, fears and wishes or unusual acts of kindness and generosity. Otherwise, the disinhibition effect may conduct to rude language, harsh criticisms, anger, hatred, and even threats.

Personal Language

McMenamin, Dongdoo Choi and Coulthard [6, 15] studied for the first time the personal language to confirm the authorship of written texts in forensics. They revealed that each person uses a particular set of meanings of words and expressions depending of education, culture and personality and gender. Even if any speaker/writer can use any word at any time, they tend to make unique co-selections of preferred words, phrases or idioms, grammar constructions. This personal vocabulary is called idiolect.

The Flaming Phenomena

A narrow distinction exists between risky topics (that could easily lead to flames) and flames (the actual aggressive interactions). The psychology of this phenomena helps identify the most important features that characterize each discussion type.

Pazienza, Stellato and Tudorache [19] showed that *flames* are a sequence of "non constructive", aggressive posts, that have no positive contribution to the discussion. In flames users attack each other at a personal level instead of contrasting the discussion partner for his/her approach, ideas, contribution or argumentation. Flames often induce moderators to close discussions. Sometimes the moderator himself generates flames due to his tough policy or offtopic interventions.

A risky discussion is the introduction for a flame and contains both flame and no-flame elements. Therefore risky topics could neither be categorized as flames, nor as normal discussion. In real life the distinction between flames and risky discussions is very subjective. It depends on moderator skills, his attitude, level of stress and level of implication in the subject. There are mini-flames (two-three flame posts) that often extinguish by themselves or at the correct intervention of a moderator. As in flames, there are two or maximum three actors (users) participating to arguments. In risky discussions more than one mini-flame could be found and moderators usually don't make hush interventions. Also in such threads users tend to get stormy about ideas and not persons as in flames. When a risky topic turns personal flaming occurs. Moreover, in risky discussions often a good number of off-topic posts are found and the evolution of the thread depends directly of moderator skills and in some measure of users' willingness to argue.

A no-flame ("normal") topic is a discussion that continues without the risk flames. It is usually a discussion in which moderators need not to interrupt discussion or recall users. At the most they act as normal users participating in the discussion.

To provide a flavour of what a flame is we show some examples of the English translation of the Italian corpus. The translation was done trying to preserve as much as possible of the original sense and style to better illustrate emotion involved. The names of people, companies and users were blinded for privacy reasons.

> ***User1** – my answers are on the other general topic which has been closed... but is correct to open another topic similar with the closed one??'*
>
> ***Moderator1** – the posts have been moved. And for certain questions there are the pm. And because you know how to use them. just use them!*
>
> ***User1** – GOOD If I'm the one creating problems I'm going to autobann myself for a few days so everybody should be happy to externalize your legal thoughts.... Thank you all... (Extract from flame 381)*

In this fragment User1 tries to get polemic with Moderator 1 and protests for the closure of his topic. It is a direct personal attack against the moderator and moderation policy. Also the moderator replies in a non professional and aggressive fashion. In a normal discussion User1 would have asked politely why his topic has been closed or better he should send a private message to a moderator as suggested.

> A risky topic English translation example follows: ***User2** - .... Just heard on the news.... The dwarf «there was no Bulgarian edict, mine was just an appeal to the new leaders coming into Company1... that» certain things «shouldn't happen any more» I wonder if there is a limit to stupidity and impudence of that little man! (Extract from risky sequence 705)*

Even if the fragment is not a flame we can anticipate that it will develop into one if not moderated properly because is a direct attack to a person with many supporters. When discussion gets more personal, or the topic is a hot topic for a certain user it could easily evolve into flame.

Flames and risky topics share many features and mostly is a subjective classification. Henceforth, it is an example of the English translation of a normal discussion. Even if the discussion topic is still hot and could easily evolve in a risky or even flame thread the dialogue partners had the right approach and attacked neither ideas nor other users. It is a bitter text but without any flaming elements: no personal attack or cites.

> ***User3** - Person2, the wife of the mafia head Person3, has cited for damages the authors of Company2 fiction 'Film1'. (...) the show, that presented the*

ascend of the boss Person3, has damaged his public image. (...) In what country are we living?
***User4** - In the pulcinella country.*
***User1** - It is a collateral negative effect of the democracy, absolutely shame...*
***User5** - I don't know, the only thing I could think of is a fruit tree with sick branches where proliferates a certain epidemic illness. (No-flame 754).*

Flames and Risky Topics Features

Different motifs could determine normal discussions to become flames. We grouped these features in two main categories: expression and user profiling. Expression is a meta-feature including the elements that show how concepts and sentiments are expressed in forums as: language, emoticons, and punctuation marks. User profiling captures the behaviour of dialogue actors including type of user, personal background, external causes and moderation policy.

Expression Features

Expression includes features as: discussion topic, general language, personal language, cites, and emoticons.

Topic: The preliminary analysis of several forums show that topics as politics, sport, social integration and even brands and working philosophies as open source versus paid software usually degenerate in flames. It is the aspect of appartenance to a group or ideology that makes people fight for. Topics like small chat or general discussions, that are not very important for the people participating, are less prone to become flames, unless a troll makes a target of that forum and attacks different topics.

Language: is a complex phenomenon and hard to model. Dialogue in open communities is even harder to model and analyse not only for the differences in language and users background but also for the non linear dynamics of web communities. (see before)

Bucci and Maskit [5] demonstrated, in their psychological clinical studies, that ambiguous phrasal constructs, mostly verbal, missing of argumentation are a sign of tension. Also the avoidance of taking responsibility over the expressed ideas introduce flames. E.g. “un” against “il”; “bene”, “niente” – Italian, “the” against “this”, “the idea” against “my idea“- English. Moreover, Italian forums are characterized by a high context communication. Users addresses directly to each other at a personal level (“cut and thrust” - “botta e risposta” in Italian). Furthermore, Culpeper demonstrates that users and are not likely to express clearly their ideas.[7] In our corpus around 18% of flames present this type of interaction.

Furthermore, in Italian flames often phrases miss the subject and express a general disagreement. (E.g. “è una cosa stupida; non è vero” - “it is a stupid idea; it is not true”). Also the use of hypocritical politeness in expressions as: “perdonami se...”, “scusa se...” (“excuse me if...”). This kind of expression is present in all the flames to which take part women and in almost 30% of men flames.

Culpeper [7] shows that most of the time offensive language generate flames and is also part of flaming. In the analyzed corpus 40% of men use offensive language, while women use little or no offensive language preferring irony. Specific symbols and marks also express emotionality as: repeated question or exclamations marks, question marks followed by exclamation marks or the written expression of emoticons like ghghg or lol. Also uppercase is used to get the attention or to underline an idea. But in the analyzed forum uppercase is seldom used. For that reason we excluded this feature form analysis.

Cites: Often, when arguing, people tend to repeat parts of the discourse of the contender. In forums such phenomena is even more present in the form of cites and cross-citing.

Emoticons: integrate the language and act as visual cues for expressing emotion as: anger, happiness, irony.

External causes: Furthermore, unknown external causes not related to the forum can lead to flames as: personal problems or simply a user having a bad day.

User Profiling

User profiling meta-feature includes several individual features as: actors (historic enemies, newbies) personal background and moderation.

Actors: Discussions usually take place between two or maximum three users on a specific topic and often the same users argue over several topics while posts are mostly verbal and missing of argumentation.

Historic enemies: There are historic enemies: pairs or small numbers of persons that for different reasons have the necessity to argue. Bucci, W. and Maskit show that it is natural to try to contrast your opponents over several topics.[5] Therefore in a certain period we will find that flames are provoked by a small group of users.

Newbies: Newly registered persons may cause flames simply because they don't know the rules of that specific community or fail to communicate with previously unknown persons.

Personal background: Each user influences directly the development of a flame. Background and education are directly reflected in the way in which a person expresses and therefore in the likelihood of provoking flames.

Moderation: Flaming occurs more frequently in forums where the administration uses a tough or a nonlinear moderation policy. Moderators can generate flames with nonsense and off topic interventions in no flame or risky discussions. Also discussions where moderators make most interventions usually are hot discussions.

Our Approach to Flames and Risky Topics Identification

In this section our approach will be presented starting with information about the corpus structure. Furthermore, the proposed model will be widely discussed including algorithm, software, parameterization and experimental setup.

Corpus

The corpus used for this experiment is a real life Italian corpus directly extracted from the politics section of a generic forum that hosts also social, sport and general topics. The corpus is composed of a 1540 posts, from which 170 flames, 330 risky posts 1040 no flames. All discussions were selected in the order of creation. The posts were extracted from 195 topics and were written by 73 different users, from which 11 females and 62 males.

Architecture

To model the flaming phenomena, two main meta-categories of features will be analyzed: expression and user profiling. The architecture consists of two main steps: corpus preprocessing and classification (performed for each experiment). (see Figure 1) Corpus preprocessing has four main purposes: spider the Web to gather the corpus, parse the gathered HTML pages to extract the text and features, manual annotation of the extracted features, features selection and feature vectors building for each posts sequence.

Classification step was performed using Support Vector Machines (SVM). Flames and risky situations develop over several posts. In analyzed forums a 5 post sequence is the range in which usually discussions go hot and out of control. Therefore, we selected as analysis unit or document (as in information retrieval) a window of 5 consecutive posts. (E.g. Posts 1-5, posts 2-6 from discussion 1).

Each analysis unit will be represented as a vector of features including expression and user profiling as presented in Table 1.

Documents	Expression Vector	User Profiling Vector
Doc 1	$E_1,..., E_k , E_n$	$UP_1,...,UP_k,UP_u$
...	...	...
Doc d	$E_1,..., E_k , E_n$	$UP_1,...,UP_k,UP_u$

Table 1 - Document representation matrix

E_k = expression features for document d
UP_u = user u profiling for document d

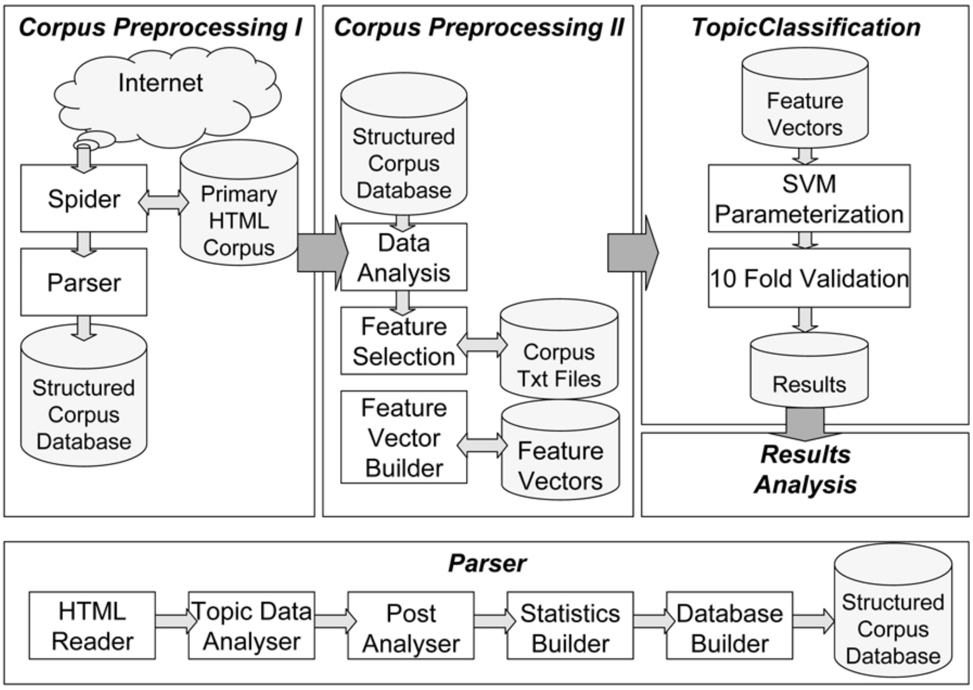

Figure 1 - Flames and Risky Topics Recognition Architecture

Corpus Preprocessing

The first part of corpus preprocessing is dedicated to collect and store the information, the second part describes corpus annotation, while the third part details data representation (feature selection and vectors builder).

Information collection: A Firefox extension was utilized to **spider** and download the contents of the politics section of the forum. To eliminate double posts and to extract features a php **parser** was developed and a MYSQL support database was used to store the processed corpus. Posts were extracted individually, together with several features as: topic, post author, user type (mod, admin or user), date and time of posting, quotes and other users' citations. Each emoticon and special punctuation mark was substituted by its class in the corpus.

Manual Post Annotation: Each post was manually annotated by two different annotators and assigned to one of the three classes: flame, risky or no-flame. The agreement rate was: 68 (Cohen's kappa coefficient with 95% confidence interval). This is considered a strong agreement as shown by Gwet. [10] A third annotator selected the final class from the initial annotations.

Feature Selection: Previously we have identified several psycho/cognitive/linguistic characteristics of the flaming phenomena.

Expression Feature Selection: In the mathematical model we will include the following expression features: authors, post contents, quotes and their authors, emoticons and special marks. Since the expression is considered as an extended specific vocabulary it will be represented by the word vector model and the features will be represented as Tf*idf, term frequency - inverse document frequency. As suggested by Lungu, Pazienza and Tudorache [12] several strategies are employed to normalize expression features. Post contents were preprocessed to eliminate the majority of sparse and frequent words. This helps to reduce the presence of misspelled words, very frequent in forums and common words that don't carry relevant semantic information. (E.g misspelled words as: "inutilite" instead of "inutile" - *useless* or "avrebe" instead of "avrebbe" - *could have*, and common words as: "ora", *now*, "perche" - *why*) A previous experiment on the same forum suggested that it is better to select words present in at least 3 documents and maximum 50. Basili and Moschitti [2] suggest eliminating words having less than 3 characters that do not carry semantics as: articles or prepositions. An Italian language stemmer was used to collapse words to their root. Words sharing the same root have similar semantics even if derive from different parts of speech. (E.g. "elettorale" – *electoral*, "elettore" - *voting person*, "eletto" - *voted*).

Then, emoticons were extracted and assigned to 8 different classes: positive, negative, angry, sad, bitter, holiday, fear and ironic as were categorized on the analyzed forum. For our experiment, emoticons were substituted by their class name. Specific symbols were also substituted by conventional classes as: irony (?!/!?), multiplequestion (???), singlequestion (?), multipleexclamation (!!!), singleexclamation (!), laugh (ghgh, lol, ahah).

User Profiling Feature Selection

For each analyzed post window, user profiling is represented as a vector of user activity features. For each user we compute three activity indicators representing his contribution to each class: flames, risky topics and no flames. For each class the activity is computed as the overall user activity combined with the local user activity (see equation 1). To minimize errors we considered that the general user activity is 0 if the user have not a clear tendency towards that class (is less than two times the average of the entire forum).

General impact of user activity (see equation 2) is viewed as a normalized weighted measure of the activity for each class in the training corpus.

Flaming index is computed as the estimate of likelihood for a given user to participate to each class, while user interventions weight is a measure of user impact on the general activity of the analysed forum.

Local user activity (see equation 3) is the number of posts written by user *u* in the analysis unit (5 posts window) and represents the contribution of user *u* to the analysed document (*d*). To give the chance of a fair analysis for newbies, inactive and unregistered users, all measures are averaged by the number of days in which the analyzed user is active.

$$\mathit{UserProfil\ ing} = \mathit{GeneralImp\ actOfUserA\ ctivity} * \mathit{LocalUserA\ ctivity} \qquad (1)$$

$$\mathit{GeneralImpactOfUserActivity} = \mathit{UserInterventionsWeigh} * \log \frac{1}{\mathit{ClassIndex}} \qquad (2)$$

$$\mathit{ClassIndex}_u = \frac{\mathit{PostsNoWrittenByUser}_u \in \mathit{Class}}{\mathit{PostsNoWrittenByUser}_u} \qquad (3)$$

$$\mathit{LocalUserActivity} = \frac{\mathit{NoOfPostsWrittenByUserU} \in d}{\mathit{NoOfAnalysedPosts}} \qquad (4)$$

Feature Vectors Builder

Once extracted each post data the feature vectors including expression and user profiling were build for each 5 posts window. The vectors were ordered first by topic and afterwards by post order following the natural order in the analyzed forum. For this step we used the “Waikato Environment for Knowledge Analysis”, Weka [28, 29] with an extension of Word Vector Tool [30] and customized software for building the feature vectors.

To simulate the choice of a human moderator, the class assignment for each post window was done empirically combining the individual post assignment and will be further considered as a heuristic. A flame post was marked with a 2 score, while risky posts with 1.8 and normal posts with a score of 1. To maintain a cautious attitude we considered risky posts nearly as dangerous as flames. For each window the each category scores are summed and the biggest one defines the window category.

Classification Algorithm

The actual classification was performed with parameterized Support Vector Machines and using stratified 10 folds cross validation model to confirm results. SVM algorithm developed by Vapnik [27] is considered currently a state-of-the-art classification algorithm. SVM offers a good generalization and can manage noisy training data. [4] For this study was adopted the Joachims' SVM implementation - SVMlight [11] with a linear kernel.

For each discussion will be computed a score that shows only a ranking of classification. Positive scores indicate that the document is classified as positive class, while negative score indicate that the document belongs to negative/neutral class. For this experiment we parameterized the cost parameter of SVM to optimize and balance both precision and recall. Since topic analysis is subjective and also prone to errors the goal was to avoid overfitting. C represents the trade off between allowing training errors and forcing rigid margins of classification. Basili, R. and Moschitti [2, 3] showed that increasing the value of C increases the cost of misclassifying points and forces the creation of a more accurate model that may not generalize well. Therefore we were looking for the smallest C and for the break-even point of Precision and Recall (intersection of Precision and Recall curves).

Evaluation Criteria

As described before the aim is to classify flames, risky topics separately and then a mixed class that contains both flames and risky topics. The no-flame class is considered neutral. For evaluation purposes, as in information retrieval, the F1 measure will be used (see equation 6), that is the weighted harmonic average of Precision and Recall measures (see equations 7, 8). Also accuracy will be calculated (see equation 6).

$$F1 = \frac{2 \times \Pr ecision \times \operatorname{Re} call}{\Pr ecision + \operatorname{Re} call} \quad \textbf{(5)} \qquad Accuracy = \frac{tp + tn}{tp + fp + tn + fn} \quad \textbf{(6)}$$

Where:

$$\Pr ecision = \frac{tp}{tp + fp} \quad \textbf{(7)} \qquad \operatorname{Re} call = \frac{tp}{tp + fn} \quad \textbf{(8)}$$

Where: tp (true positives) - topics correctly identified;
fn (false negatives) = not found correct topics;
fp (false positives) = incorrect topics marked as positives.

Results

In this section we present the results of flame, risky topics and mixed (flame and risky topic) identification against no-flames that acts as neutral class.

A second group of experiments was conducted to classify no-flames against the other classes: flames, risky topics and mixed class (flames and risky topics). Other two experiments were conducted to study the differences between flames and risky topics. For each experiment the expression features and the combination between expression features and user profiling results was evaluated.

Experiment	***Flames vs. No-flames***		***Risky Topics vs. No-flames***		***Mixed Class vs. No-flames***	
Results	**F1**	**Acc(%)**	**F1**	**Acc (%)**	**F1**	**Acc.(%)**
Fold 1	90.00	92.31	82.76	84.21	87.84	85.37
Fold 2	83.64	88.46	84.78	85.26	89.61	86.99
Fold 3	88.89	92.31	88.64	89.47	87.50	85.25

Fold 4	89.29	92.31	75.61	78.95	83.92	81.15
Fold 5	90.19	93.59	85.06	86.32	90.41	88.52
Fold 6	87.27	90.91	77.78	78.95	85.91	82.79
Fold 7	94.34	96.10	75.79	75.79	86.25	81.97
Fold 8	92.59	94.81	87.36	88.42	91.67	90.16
Fold 9	92.59	94.74	81.82	82.80	87.84	85.12
Fold 10	94.12	95.95	88.89	90.11	88.57	86.55
Average	**90.29**	**93.15**	**82.85**	**84.03**	**87.95**	**85.39**

Table 2 - Flames and Risky Topics Mixed Class Recognition, F1 score and Accuracy

Experiment	***No-flames vs. Flames***		***No-flames vs. Risky Topics***		***No-flames vs. Mixed Class***	
Results	**F1**	**Acc(%)**	**F1**	**Acc(%)**	**F1**	**Acc(%)**
Fold 1	93.75	92.31	85.43	84.21	87.84	85.37
Fold 2	91.09	88.46	85.71	85.26	89.62	86.99
Fold 3	94.12	92.31	91.20	89.47	87.50	85.25
Fold 4	94.00	92.31	81.48	78.95	83.92	81.15
Fold 5	95.24	93.56	87.38	86.32	90.42	88.52
Fold 6	92.94	90.91	80.00	78.95	85.90	82.79
Fold 7	97.03	96.10	75.79	75.95	86.25	81.97
Fold 8	96.00	94.81	89.32	88.42	91.67	90.16
Fold 9	95.92	94.74	83.67	82.80	87.84	85.12
Fold 10	96.91	95.95	91.09	90.11	88.57	86.55
Average	**94.70**	**93.15**	**85.11**	**84.04**	**87.95**	**85.39**

Table 3 – No-flames Recognition, F1 score and Accuracy

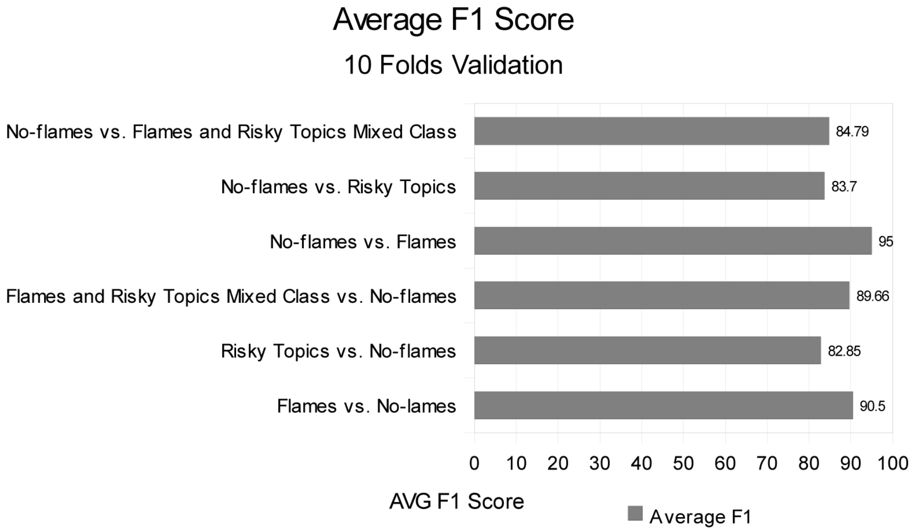

Figure 2 - Flames and Risky Topics Recognition Average F1 Score for 10 Fold Validation

For our experiments C parameter was computed in 0.1 steps and Precision and Recall converged for C=1. User profiling information improved the expression results in average by 0.2 (F1 score). Partial results will not be reported.

In Table 2 are showed the final results (F1 score and Accuracy) of the experiments for flames, risky topics and mixed class recognition for C=1, while in Table 3 are analyzed the results of no-flames classification. As shown in Figure 2 for risky topics against flames classification the average F1 score was 91.17 and the average accuracy 88.92%.

Conclusions and Future Work

Situated at the confluence of many different disciplines: computational linguistics, psychology and social anthropology, this study has a great impact not only for identifying flames but also for understanding the dynamics of web forums. The study could help create a better web environment by identifying as early as possible risky topics and flames as well as hot topics. Several areas of application were identified as: post author identification over different forums, troll identification, email flame filtering, and threats management (identify angry, subversive persons or groups of persons in political or corporate forums and social networks). Among others, immediate commercial applications of our approach could be in market research field as, for example, to develop opinion mining applications about products, brands and companies.

Several aspects were studied as: psychology of web communication (forums as social groups, communication barriers, the lack of non verbal for expressing emotion), language correlated problems (mother language, personal language, misspelled words, improper grammar usage), user behaviour (determined by the interaction between users, the presence of new users, historic enemies or friends, background and gender).

Experiments confirmed in our case that language has a major influence in flames recognition, while user behaviour had only a minimal influence on results. We obtained results comparable with the state of the art. Our average accuracy is 83% for risky topics recognition, 85% for mixed class recognition, 88.92% for risky topics vs. flames recognition and 93% for flames recognition, while for example the best accuracy obtained by Pang and Lee [18] for sentiment classification is 83% for unigrams. Shi, L. [24] obtained 87% accuracy for web forum sentiment analysis based on topics, while Gupta et al. [9] obtained 83% accuracy and 0.72 best F1 score for customer care emotional emails recognition. Xu and Zhu [31] obtained 90.94% accuracy for filtering offensive language in online communities using grammatical relations, while Ptaszynski et al. [21] obtained F1 – score of 88.2% for affect analysis against cyber-bullying.

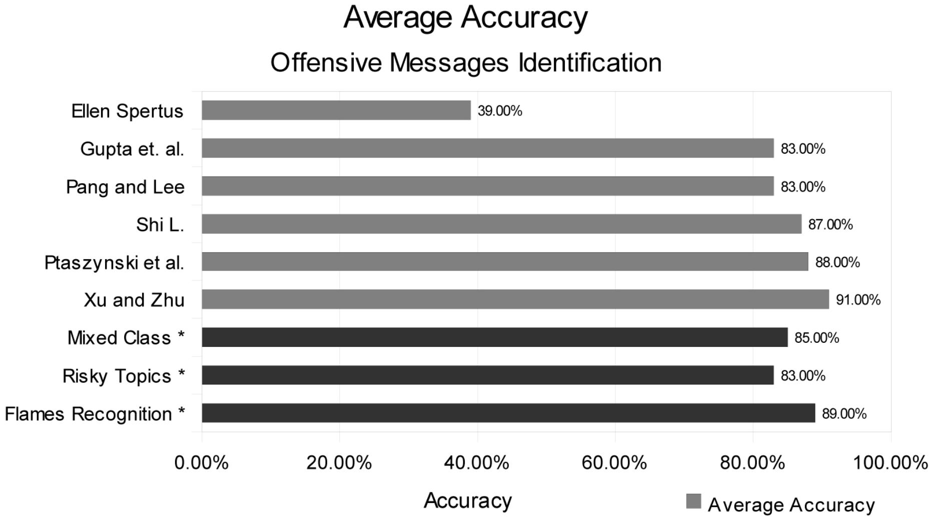

Figure 3 - Flames and Risky Topics Recognition Comparative Average Accuracy

Earlier Ellen Spertus' Smokey (based in C 4.5 algorithm) obtained a accuracy of 39% for flames and 97% on no-flames class. [25] Results show that it is difficult to obtain an optimal recognition on the closest classes as: flames and risky topics and no flames and risky topics. Furthermore the differences between the folds results confirm that the annotation process itself is subjective and varies even for the same annotator depending on his mood, level of attention and stress. This indicates that results could be improved by better annotation methodologies.

Further research should be necessary to refine the user behaviour model. One study direction could be the analysis of the variance of behaviour for each user during flames, risky topics and normal discussions.

Expression analysis could be improved using latent semantic analysis to discover different associations of concepts that determine flames, as well as word sense disambiguation and name entity recognition to identify for example the nicknames of different persons referred in discussions. Moreover, other classifiers (lexicon-based) could be tested as suggested by the work of Georgios Paltoglou et. Al. [17] that focused on analyzing the emotionality involved in social media responses. Furthermore, a NER algorithm could be used to identify the hot topics regarding products and brands. The opinions about direct competitor brands and products could be analyzed and evaluated in terms of customer satisfaction. In the same context the market share of a product could be evaluated computing how many users (and their profile) that had already purchased the product and the ones willing to. This information could be integrated into a business portal as suggested by Lungu, Velicanu, Bara and Botha [13].

Moreover, as Basili R., Moschitti, A., Pazienza and Zanzotto suggest-personalizing the presentation pages of products using information extraction algorithms able identify the hottest features of products discussed in forums could further improve the market share and income of a company. [4]

REFERENCES

[1]**Adler, R. B., Rodman, G., and Hutchinson, C. (2011),** ***Understanding Human Communication***. Oxford University Press;

[2]**Basili, R. and Moschitti, A. (2005),** ***Automatic Text Categorization: From Information Retrieval to Support Vector Learning***. Aracne Editrice, Informatica;

[3]**Basili, R. (2003),** ***Review of «Learning to Classify Text Using Support Vector Machines» by Thorsten Joachims***, Computational Linguistics, 29, 4, pages. 655-661;

[4]**Basili R., Moschitti, A., Pazienza and Zanzotto, F.M. (2003),** ***Personalizing Web Publishing via Information Extraction***. *IEEE Journal of Intelligent Systems*, Jan/Feb 2003, IEEE Computer Society;

[5]**Bucci, W. and Maskit, B. (2006),** ***A Weighted Dictionary for Referential Activity. Computing Attitude and Affect in Text;*** pp. 49-60. Springer Netherlands. The Information Retrieval Series;

[6]**Coulthard, M. (2004),** ***Author Identification, Idiolect, and Linguistic Uniqueness: Forensic Linguistics***. Vol. 4 no. 25, pp. 431-447. Oxford University Press;

[7]**Culpeper, J. (2011),** ***Impoliteness: Using Language to Cause Offence***. Cambridge University Press;

[8]**Gudykunst, W. B. (2005),** ***Theorizing about Intercultural Communication***. Thousand Oaks: Sage;

[9]**Gupta, N., Gilbert, M., and Di Fabbrizio, G. (2010),** ***Emotion Detection in Email Customer Care***, CAAGET '10 pp. 10-16. Los Angeles, California, US: Association for Computational Linguistics;

[10]**Gwet, K. (2010),** ***Handbook of Inter-Rater Reliability***. STATAXIS Publishing Company;

[11]**Joachims, T. (2002),** ***Learning to Classify Text Using Support Vector Machines.*** Kluwer Academic Publishers. Kluwer international series in engineering and computer science;

[12]**Lungu, I., Pazienza, M. T., and Tudorache, A. (2009),** ***Organic Topic Recognition in Online Documents.*** *Economic Computation and Economic Cybernetics Studies and Research,* 43 (4), pp. 73-85, 2009;

[13]**Lungu, I., Velicanu, M. T., Bara, A. and Botha, I. (2009),** ***Portal Based System Integration – Foundation for Decision Support.*** *Economic Computation*

and Economic Cybernetics Studies and Research, 43 (1), pp. 123-134, 2009;
[14]**Mabry, E. A. (1997),** ***Framing Flames: The Structure of Argumentative Messages on the Net.*** *Journal of Computer-Mediated Communication*, 2(4);
[15]**McMenamin, G. R. and Dongdoo Choi (2002),** ***Forensic Linguistics: Advances in Forensic Stylistics*** . London: CRC Press;
[16]**Mehrabian, A. (1971),** ***Silent Messages***. Wadsworth Publishing Company;
[17]**Paltoglou, G. et al. (2010),** ***Sentiment Analysis of Informal Textual Communication in Cyberspace***. In *Proceedings of Engage 2010*, London, UK;
[18]**Pang, B. and Lee, L. (2008),** ***Opinion Mining and Sentiment Analysis***. Foundations and Trends in Information Retrieval vol. Vol. 2, no. 1,2, pp. 1-135;
[19]**Pazienza, M. T., Stellato, A., and Tudorache, A. (2008),** ***Flame, Risky Discussions, No Flames Recognition in Forums***. In *Proceedings of Sentiment Analysis: Emotion, Metaphor, Ontology and Terminology, EMOT 2008*, Marrakesh, Morocco;
[20]**Peck, M. S. (1987),** ***The Different Drum: Community Making and Peace***. New York: Simon & Shuster;
[21]**Ptaszynski, M. et al. (2010),** ***Machine Learning and Affect Analysis against Cyber-Bullying***. In *Proceedings of AISB 2010*, Leicester, UK;
[22]**Riordan, M. A. and Kreuz, R. J. (2010),** ***Cues in Computer-mediated Communication: A Corpus Analysis***. *Journal of Computers in Human Behavior*,(26): pp. 1806-1817;
[23]**Shannon, C. E. and Weaver, W. (1949),** ***The Mathematical Theory of Communication***. Urbana, Illinois: University of Illinois Press;
[24]**Shi, L. et al. (2009),** ***Web Forum Sentiment Analysis Based on Topics***. In *Proceedings of Ninth IEEE International Conference on Computer and Information Technology 2009*, IEEE Computer Society;
[25]**Spertus, E. (1997),** ***Smokey: Automatic Recognition of Hostile Messages***. *In Proceedings of IAAI*, pp. 1058—1065;
[26]**Suler, J. (2008),** ***The Basic Psychological Features of Cyberspace***. The Psychology of Cyberspace;
[27]**Vapnik, V. N. (1995),** ***The Nature of Statistical Learning Theory***. Springer-Verlag New York, Inc.;
[28]**Weka (2010),** ***Weka 3: Data Mining Software in Java.***
[29]**Witten, I. H. and Frank, E. (2005),** ***Data Mining: Practical Machine Learning Tools and Techniques***. Vol. Second Edition. Morgan Kaufmann. Data Management Systems;
[30]**Word Vector Tools (2010),** ***The Word & Web Vector Tool.***
[31]**Xu, Z. and Zhu, S. (2010),** ***Filtering Offensive Language in Online Communities using Grammatical Relations***. In *Proceedings of Collaboration, Electronic messaging, Anti-Abuse and Spam Conference 2010*, Redmond, Washington, US.