International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 6 ǁ 2022 ǁ PP. 672-677

Enhanced Message Filtering Classification to Filter Profanity Words from Online Social Networks

G. RAVINDRA BABU[1], M. SAI SAMYUKTHA[2], V. JYOTHY SRAVYA SRI[3],S. SANDEEP KUMAR[4], V. ANISH[5]

[1]*Assistant Professor, Dept. of CSE, Sai Spurthi Institute of Technology,Khammam,Telangana,India*
[2,3,4,5,]*B.Tech Student, Dept. of CSE, Sai Spurthi Institute of Technology, Khammam,Telangana,India*

ABSTRACT: *One key issue in the present Online Social Networks (OSNs) is to enable clients to control the messages posted on their own private space to keep away from that undesirable substance is shown. Up to now, OSNs offer little help to this prerequisite. To fill the hole, in this paper, we propose a framework permitting OSN clients to have an immediate control on the messages posted on their dividers. This is accomplished through an adaptable rule based framework, that permits clients to redo the separating rules to be applied to their dividers, and a AI based delicate classifier consequently naming messages on the side of content-based sifting.*

INDEX TERMS: *Online social networks, information filtering, short text classification, policy-based personalization*

Date of Submission: 29-05-2022 Date of acceptance: 10-06-2022

I. INTRODUCTION

Online Social Networks (OSNs) are today one of the most well known intuitive medium to impart, share, and spread a lot of human existence data. Day by day and ceaseless interchanges suggest the trading of a few kinds of content, including free text, picture, sound, and video information. As per Facebook statistics1 normal client makes 90 bits of content each month, though in excess of 30 billion bits of content (web joins, reports, blog entries, notes, photograph collections, and so on) are shared every month. The enormous and dynamic person of these information makes the reason for the work of web content mining systems meant to naturally find helpful data torpid inside the information. They are instrumental to offer a functioning help in complex and modern undertakings engaged with OSN the executives, for example, for example access control or data separating. Data separating has been significantly investigated for what concerns printed records and, all the more as of late, web content. Nonetheless, the point of most of these proposition is essentially to give clients a characterization instrument to keep away from they are overpowered by futile information. In OSNs, data sifting can likewise be utilized for a unique, more touchy, reason. This is because of the reality that in OSNs there is the chance of posting or remarking different posts on specific public/private regions, brought in everyday dividers. Data separating can along these lines be used to enable clients to consequently control the messages composed on their own dividers, by sifting through undesirable messages. We accept that this is a vital OSN administration that has not been given up to this point. Without a doubt, today OSNs offer next to no help to forestall undesirable messages on client dividers. For instance, Facebook permits clients to state who is permitted to embed messages in their dividers (i.e., companions, companions of companions, or characterized gatherings of companions). Be that as it may, no substance based inclinations are upheld and along these lines it is unimaginable to expect to forestall undesired messages, for example, political or profane ones, regardless of the client who posts them. Offering this support isn't just a matter of utilizing recently characterized web content mining strategies for an alternate application, rather it needs to plan impromptu grouping techniques. This is on the grounds that divider messages are established by short message for which customary grouping strategies have genuine constraints since short texts don't give adequate word events.

The point of the current work is in this way to propose and tentatively assess a mechanized framework, called Filtered Divider (FW), ready to channel undesirable messages from OSN client dividers. We exploit Machine Learning (ML) text arrangement procedures to consequently relegate with each short text message a bunch of classifications dependent on its substance.

II. RELATED WORK

2.1 Content-Based Filtering

In content-based separating, every client is accepted to work freely. Subsequently, a substance based sifting framework chooses data things dependent on the connection between the substance of the things and the client inclinations as gone against to a cooperative sifting framework that picks things dependent on the relationship between's kin with comparable inclinations. While electronic mail was the first area of early work on data separating, ensuing papers have tended to broadened spaces including newswire articles, Internet "news" articles, and more extensive network assets. Reports handled in content-based sifting are for the most part text based in nature and this makes content-based separating near text arrangement. The movement of separating can be demonstrated, truth be told, as an instance of single mark, parallel order, apportioning approaching reports into important and no relevant classes. More complicated separating frameworks incorporate multilabel text classification consequently marking messages into halfway topical classifications. Content-put together sifting is basically based with respect to the utilization of the ML worldview as indicated by which a classifier is consequently prompted by gaining from a bunch of preclassified models. A noteworthy assortment of related work has as of late showed up, which vary for the took on include extraction techniques, model learning, and assortment of tests. The element extraction method maps message into a smaller portrayal of its content and is consistently applied to preparing and speculation stages. A few examinations demonstrate that Bag of-Words (BoW) approaches yield great execution and win in general over more modern message portrayal that may have predominant semantics however lower factual quality. All things considered, there are various significant methodologies in content-based separating and text arrangement overall appearance common benefits and detriments in capacity of application dependent issues. In, a nitty gritty examination investigation has been led affirming prevalence of Boosting-based classifiers, Neural Networks, and Support Vector Machines over other well known strategies, for example, Rocchio and Naı¨ve Bayesian. Be that as it may, it is worth to take note of that the greater part of the business related to message separating by ML has been applied for long-structure text and the evaluated execution of the text order techniques rigorously relies upon the idea of text based reports. The use of content-put together sifting with respect to messages posted on OSN client dividers represents extra difficulties given the short length of these messages other than the wide scope of points that can be examined. Short text arrangement has gotten up to now hardly any consideration in the logical local area. Ongoing work features challenges in characterizing vigorous elements, basically because of the way that the portrayal of the short text is succinct, with numerous incorrect spellings, nonstandard terms, and clamor.

2.2 Policy-Based Personalization of OSN Contents

As of late, there have been a few proposition taking advantage of grouping instruments for customizing access in OSNs. For example, in [27], a characterization strategy has been proposed to arrange short instant messages to keep away from overpowering clients of microblogging administrations by crude information. The framework depicted in [27] centers around Twitter2 and partners a bunch of classes with each tweet depicting its content. The client would then be able to see just particular kinds of tweets in light of his/her advantages. Conversely, Golbeck and Kuter[28] propose an application, called FilmTrust, that adventures OSN trust connections and provenance data to customize admittance to the site. Notwithstanding, such frameworks try not to give a separating strategy layer by which the client can take advantage of the aftereffect of the characterization cycle to choose how also to which degree sifting through undesirable data. In contrast, our separating strategy language permits the setting of FRs as indicated by an assortment of rules, that don't consider just the consequences of the grouping system yet additionally the connections of the divider proprietor with other OSN clients as well as data on the client profile. Also, our framework is supplemented by an adaptable instrument for BL the executives that gives a further chance of customization to the sifting strategy.

The main long range informal communication administration we know about giving separating capacities to its clients is MyWOT,3 a social organizing administration which enables its endorsers of: 1) rate assets concerning four measures: dependability, merchant unwavering quality, security, and youngster wellbeing; 2) determine inclinations deciding if the program should hinder admittance to a given asset, or ought to just return acautioning message based on the predefined rating. Regardless of the presence of certain likenesses, the methodology embraced by MyWOT is very not quite the same as our own. In specific, it upholds separating rules which are undeniably less adaptable than the ones of Filtered Wall since they are as it were in light of the four previously mentioned rules. Additionally, no programmed order instrument is given as far as possible client.

3 FILTERING RULES AND BLACKLIST MANAGEMENT

In this segment, we present the standard layer took on for sifting undesirable messages. We start by depicting FRs, then, at that point, we represent the utilization of BLs.

3.1 Filtering Rules

In characterizing the language for FRs particular, we consider three primary issues that, as we would see it, should influence a message separating choice. As a matter of first importance, in OSNs like in daily existence, a similar message might have unique implications and importance dependent on who composes it. As a result, FRs ought to permit clients to state limitations on message makers. Makers on which a FR applies can be chosen based on a few unique models, one of the most pertinent is by forcing conditions on their profile's credits. In such a manner it is, for example, conceivable to characterize rules applying just to youthful makers or to makers witha given strict/political view. Given the informal community situation, makers may likewise be recognized by taking advantage of data on their social chart. This infers to state conditions on type, profundity, and trust upsides of the relationship(s) makers ought to be engaged with request to apply them the predetermined guidelines. This large number of choices are formalized by the thought of maker particular, characterized as follows:

Definition 1. A maker determination creatorSpec verifiably indicates a bunch of OSN clients. It can have one of the accompanying structures, potentially consolidated:

1. A bunch of quality requirements of the structure an OP av, where an is a client profile characteristic name, av and OP are, individually, a profile trait esteem and a correlation administrator, viable with a's space.

2. A bunch of relationship requirements of the structure ðm; rt; minDepth; maxT rustþ, meaning all the OSN clients partaking with client m in a relationship of type rt, having a profundity more noteworthy than or equivalent to minDepth, and a trust esteem not exactly or equivalent to maxT rust.

Definition 2. A separating rule FR is a tuple, where.

- creator is the client who determines the standard;.
- creatorSpec is a maker detail, indicatedas per Definition 1;.
- contentSpec is a Boolean articulation characterized on content limitations of the structure ðC; mlþ, where C is a class of the first or second level and ml is the base enrollment level limit needed for class C to make the limitation fulfilled; .
- activity €{block; notify}g means the activity to be performed by the framework on the messages coordinating contentSpec and made by clients distinguished by creatorSpec.

3.2 Online Setup Assistant for FRs Thresholds

As referenced in the past segment, we address the issue of setting edges to channel rules, by imagining what's more carrying out inside FW, an Online Setup Assistant methodology. OSA gives the client a bunch of messages chosen from the informational collection. For each message, the client advises the framework the choice to acknowledge or reject the message. The assortment and handling of client choices on a satisfactory arrangement of messages appropriated over every one of the classes permits to process modified edges addressing the client mentality in tolerating or dismissing certain substance

The above-depicted method can be deciphered as a participation work elicitation method inside the fluffy set system.

$$\phi(m_a, m_b) = \frac{1}{2} + \begin{cases} m_b/10 & \text{if } m_a = Filter \\ -m_b/10 & \text{if } m_a = Pass. \end{cases}$$

The participation an incentive for the nonneutral class C is controlled by applying the defuzzyfication method depicted to FC, this worth is then picked as a limit in characterizing the separating strategy.

3.3 Blacklists

A further part of our framework is a BL component to keep away from messages from undesired makers, free from their substance. BLs are straightforwardly overseen by the framework, which ought to have the option to figure out who are the clients to be embedded in the BL and choose when clients maintenance in the BL is done. To upgrade adaptability, such data are given to the framework through a bunch of rules, from this point forward called BL rules. Such principles are not characterized by the SNMP; hence, they are not implied as broad significant level orders to be applied to the entire local area. Rather, we choose to let the actual clients, i.e., the divider's proprietors to determine BL rules directing who must be restricted from their dividers and for how long. In this way, a client may be prohibited from a divider, by, simultaneously, having the option to post in different dividers

Definition 3. A BL rule is a tuple ðauthor, creatorSpec, creatorBehavior, Tþ, where

- creator is the OSN client who indicates the standard, i.e., the divider proprietor;
- creatorSpec is a maker determination, indicated as indicated by Definition 1;
- creatorBehavior comprises of two parts RFBlocked and minBanned.

- T means the time-frame the clients recognized bycreatorSpec and creatorBehavior must be prohibited from creator divider.

III. EVALUATION

4.1 Problem and Data Set Description

The investigation of related work has featured the absence of a openly accessible benchmark for looking at changed ways to deal with content-based grouping of OSN short texts. To adapt to this need, we have fabricated and made accessible an informational index D of messages taken from Facebook.61,000 200 and 66 messages from openly open Italian gatherings have been chosen and separated through a robotized methodology that eliminates undesired spam messages and, for each message, stores the message body and the name of the gathering from which it starts. The messages come from the gathering's site page segment, where any enrolled client can post a new message or answer to messages previously posted by different clients.

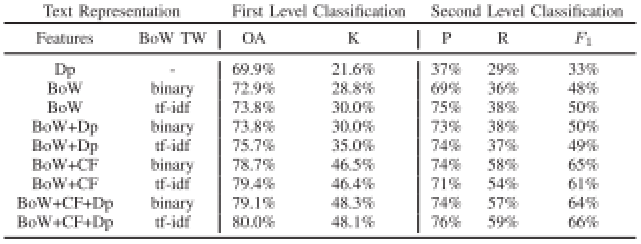
The gathering of specialists has been picked trying to guarantee high heterogeneity concerning sex, age, business, instruction, and religion. To make an agreement concerning the significance of the Neutral class and general standards in doling out multiclass enrollment we welcomed specialists to take an interest to a devoted tuning meeting.

We know about the way that the outrageous variety of OSNs content and the proceeding with advancement of correspondence styles make the need of utilizing a few informational collections as a reference benchmark. We trust that our informational index will clear the way for a quantitative and more exact examination of OSN short text grouping techniques.

4.2 Short Text Classifier Evaluation

4.2.1 Evaluation Metrics

Two distinct sorts of measures will be utilized to assess the viability of first-level and second-level characterizations. In the primary level, the short text grouping system is assessed based on the possibility table methodology. In specific, the inferred notable Overall Accuracy (OA) list catching the straightforward percent understanding between truth and arrangement results, is supplemented with the Cohen's KAPPA (K) coefficient thought to be a more hearty measure considering the arrangement happening by chance



Text Representation		First Level Classification		Second Level Classification		
Features	BoW TW	OA	K	P	R	F_1
Dp	-	69.9%	21.6%	37%	29%	33%
BoW	binary	72.9%	28.8%	69%	36%	48%
BoW	tf-idf	73.8%	30.0%	75%	38%	50%
BoW+Dp	binary	73.8%	30.0%	73%	38%	50%
BoW+Dp	tf-idf	75.7%	35.0%	74%	37%	49%
BoW+CF	binary	78.7%	46.5%	74%	58%	65%
BoW+CF	tf-idf	79.4%	46.4%	71%	54%	61%
BoW+CF+Dp	binary	79.1%	48.3%	74%	57%	64%
BoW+CF+Dp	tf-idf	80.0%	48.1%	76%	59%	66%

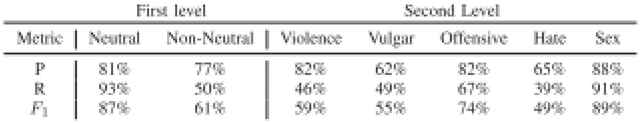
.

TABLE 1 Results for the Two Stages of the Proposed Hierarchical Classifier

4.2.2 Numerical Results

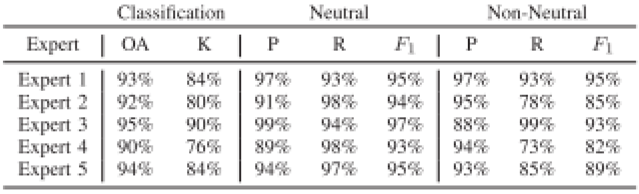
By experimentation, we tracked down a very decent boundary setup for the RBFN learning model. The best worth for the M boundary, that decides the quantity of Basis Work, is heuristically addressed to N/2, where N is the number of information designs from the informational collection. The worth utilized for the spread σ, which normally relies upon the information, is σ = 32 for the two organizations M1 and M2. As, the message has been addressed with the BoW highlight model along with a bunch of extra elements Dp what's more context oriented highlights. To work out Correct words and Bad words Dp highlights, we utilized two explicit Italian word-records, one of these is the CoLFIS corpus. The cardinalities of TrS_D and TeS_D, subsets of D with $TrS_D \cap TeS_D = ø$;, were picked so TrS_D is two times bigger than TeS_D.

Network M1 has been assessed utilizing the OA and the K worth. Accuracy, Recall, and F-Measure were utilized for the M2 network in light of the fact that, in this specific case, each example can be doled out to at least one classes.



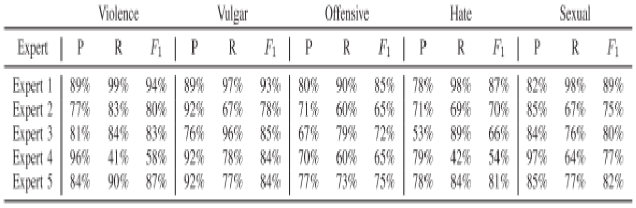
	First level		Second Level				
Metric	Neutral	Non-Neutral	Violence	Vulgar	Offensive	Hate	Sex
P	81%	77%	82%	62%	82%	65%	88%
R	93%	50%	46%	49%	67%	39%	91%
F_1	87%	61%	59%	55%	74%	49%	89%

TABLE 2 Results of the Proposed Model in Term of Precision (P), Recall (R), and F-Measure ðF1Þ Values for Each Class



	Classification		Neutral			Non-Neutral		
Expert	OA	K	P	R	F_1	P	R	F_1
Expert 1	93%	84%	97%	93%	95%	97%	93%	95%
Expert 2	92%	80%	91%	98%	94%	95%	78%	85%
Expert 3	95%	90%	99%	94%	97%	88%	99%	93%
Expert 4	90%	76%	89%	98%	93%	94%	73%	82%
Expert 5	94%	84%	94%	97%	95%	93%	85%	89%

TABLE 3 Agreement between Five Experts on Message Neutrality



	Violence			Vulgar			Offensive			Hate			Sexual		
Expert	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Expert 1	89%	99%	94%	89%	97%	93%	80%	90%	85%	78%	98%	87%	82%	98%	89%
Expert 2	77%	83%	80%	92%	67%	78%	71%	60%	65%	71%	69%	70%	85%	67%	75%
Expert 3	81%	84%	83%	76%	96%	85%	67%	79%	72%	53%	89%	66%	84%	76%	80%
Expert 4	96%	41%	58%	92%	78%	84%	70%	60%	65%	79%	42%	54%	97%	64%	77%
Expert 5	84%	90%	87%	92%	77%	84%	77%	73%	75%	78%	84%	81%	85%	77%	82%

TABLE 4 Agreement between Five Experts on Nonneutral Classes Identification

4.2.3 Comparison Analysis

The absence of benchmarks for OSN short text grouping makes dangerous the improvement of a dependable near investigation. Notwithstanding, a backhanded examination of our technique should be possible with work that show similitudes or reciprocal viewpoints with our answer. A review that reacts to these qualities is proposed in [27], where a grouping of approaching tweets into five classifications is depicted. Likewise to our methodology, messages are very short and addressed in the learning structure with both interior, content-based and context oriented properties. Specifically, the elements considered in [27] are BoW, Author Name, in addition to eight archive properties highlights.

4.3 Overall Performance and Discussion

To give a general evaluation of how successfully the framework applies a FR. This table
permits us to appraise the Precision and Recall of our FRs, since values announced in Table 2 have been figured for FRs with content detail part set to (C; 0:5), where C€Ω . Allow us to assume that the framework applies a given rule on a specific message. Thusly, Precision announced is the likelihood that the choice taken on the considered message (that is, hindering it or not) is really the right one. Conversely, Recall must be deciphered as the likelihood that, given a standard that should be applied over a certain message, the standard is truly upheld. Allow us now to examine, with a few models, the outcomes, which reports Precision and Recall esteems. The second segment of addresses the Precision and the Recall esteem processed for FRs with (Neutral; 0:5) substance requirement. In contrast, the fifth segment stores the Precision and the Recall esteem processed for FRs with (V ulgar; 0:5) imperative. Results accomplished by the substance based detail part, on the principal level characterization, can be viewed as adequate and sensibly lined up with those acquired by notable data sifting strategies. Results acquired for the substance based determination part on the subsequent level are somewhat less splendid than those got for the first, however we ought to decipher this taking into account the characteristic challenges in allotting to a messages a semantically most explicit class. In any case, the investigation of the highlights announced in Table 1 shows that the presentation of logical data (CF) fundamentally works on the capacity of the classifier to accurately recognize nonneutral classes. This outcome makes more dependable all arrangements taking advantage of nonneutral classes, which are the larger part in true situations.

IV. CONCLUSIONS

In this paper, we have introduced a framework to channel undesired messages from OSN dividers. The framework takes advantage of a ML delicate classifier to authorize adjustable substance subordinate Frs. Moreover, the adaptability of the framework as far as separating choices is improved through the administration of BLs.

This work is the initial step of a more extensive venture. The early empowering results we have acquired on the arrangement system brief us to proceed with other work that will intend to work on the nature of grouping. Specifically, likely arrangements examine a more profound examination on two associated errands. The main worries the extraction and or then again choice of logical highlights that have been displayed to have a high discriminative power. The subsequent assignment includes the learning stage. Since the fundamental area is powerfully changing, the assortment of preclassified information may not be delegate in the more drawn out term. The present group learning technique, in view of the fundamental assortment of the whole arrangement of named information from specialists, permitted an precise exploratory assessment yet should be advanced to incorporate new functional necessities. In future work, we plan to resolve this issue by exploring the utilization of internet learning standards ready to incorporate name inputs from clients. Furthermore, we intend to improve our framework with a more modern way to deal with choose when a client ought to be embedded into a BL.

The advancement of a GUI and a bunch of related devices to make simpler BL and FR detail is likewise a course we plan to examine, since ease of use is a vital necessity for such sort of utilizations. Specifically, we focus on exploring an instrument ready to naturally suggest trust values for those contacts client doesn't by and by known. We really do accept that such an apparatus ought to propose trust esteem in light of clients activities, practices, and notoriety in OSN, which may infer to improve OSN with review instruments. Be that as it may, the plan of these review based devices is convoluted by a few issues, similar to the ramifications a review framework may have on clients security as well as the limits on what it is feasible to review in current OSNs. A fundamental work toward this path has been done in the setting of trust esteems utilized for OSN access control purposes [52]. Nonetheless, we might want to comment that the framework proposed in this paper addresses simply the center arrangement of functionalities expected to give a refined instrument to OSN message sifting. Regardless of whether we have supplemented our framework with an internet based collaborator to set FR edges, the improvement of a total framework effectively usable by normal OSN clients is a wide theme which is out of the extent of the current paper. Accordingly, the created Facebook application is to be implied as a proof-of-ideas of the framework center functionalities, rather than a completely evolved framework. Besides, we know that a usable GUI proved unable adequately be, addressing just the initial step. To be sure, the proposed framework might endure of issues like those experienced in the particular of OSN protection settings. In this specific situation, numerous exact examinations [53] have shown that normal OSN clients experience issues in seeing moreover the straightforward security settings given by today OSNs. To conquer this issue, a promising pattern is to take advantage of information mining methods to induce the best security inclinations to recommend to OSN clients, based on the accessible social network information [54]. As future work, we plan to take advantage of comparable methods to construe BL rules and FRs.

Moreover, we intend to concentrate on procedures and strategies restricting the deductions that a client can do on the upheld separating rules determined to sidestep the sifting framework, for example, for example haphazardly informing a message that ought to rather be hindered, or identifying adjustments to profile ascribes that have been made for the as it were motivation behind overcoming the separating framework.

REFERENCES

[1]. Adomavicius and G. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.

[2]. M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, "Content-Based Filtering in On-Line Social Networks," Proc. ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML '10), 2010.

[3]. S. Pollock, "A Rule-Based Message Filtering System," ACM Trans. Office Information Systems, vol. 6, no. 3, pp. 232-254, 1988.

[4]. Y. Zhang and J. Callan, "Maximum Likelihood Estimation for Filtering Thresholds," Proc. 24th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 294-302, 2001

[5]. H. Schu¨tze, D.A. Hull, and J.O. Pedersen, "A Comparison of Classifiers and Document Representations for the Routing Problem," Proc. 18th Ann. ACM/SIGIR Conf. Research and Development in Information Retrieval , pp. 229-237, 1995.

[6]. S. Zelikovitz and H. Hirsh, "Improving Short Text Classification Using Unlabeled Background Knowledge," Proc. 17th Int'l Conf. Machine Learning (ICML '00), P. Langley, ed., pp. 1183-1190, 2000.

[7]. Uszok, J.M. Bradshaw, M. Johnson, R. Jeffers, A. Tate, J. Dalton, and S. Aitken, "Kaos Policy Management for Semantic Web Services," IEEE Intelligent Systems, vol. 19, no. 4, pp. 32-41, July/ Aug. 2004.

[8]. M. Carullo, E. Binaghi, and I. Gallo, "An Online Document Clustering Technique for Short Web Contents," Pattern Recognition Letters, vol. 30, pp. 870-876, July 2009.

[9]. M.J.D. Powell, "Radial Basis Functions for Multivariable Interpolation: A Review," Algorithms for Approximation, pp. 143-167, Clarendon Press, 1987.

[10]. J.A. Golbeck, "Computing and Applying Trust in Web-Based Social Networks," PhD dissertation, Graduate School of the Univ. of Maryland, College Park, 2005.

[11]. Laudanna, A.M. Thornton, G. Brown, C. Burani, and L. Marconi, "Un Corpus Dell'Italiano Scritto Contemporaneo Dalla Parte Del Ricevente," III Giornate internazionali di Analisi Statistica dei Dati Testuali, vol. 1, pp. 103-109, 1995.

[12]. L. Fang and K. LeFevre, "Privacy Wizards for Social Networking Sites," Proc. 19th Int'l Conf. World Wide Web (WWW '10), pp. 351- 360, 2010.

[13]. M. Chau and H. Chen, "A Machine Learning Approach to Web Page Filtering Using Content and Structure Analysis," Decision Support Systems, vol. 44, no. 2, pp. 482-494, 2008.

[14]. F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[15]. N.J. Belkin and W.B. Croft, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" Comm. ACM, vol. 35, no. 12, pp. 29-38, 1992.

[16]. P.W. Foltz and S.T. Dumais, "Personalized Information Delivery: An Analysis of Information Filtering Methods," Comm. ACM, vol. 35, no. 12, pp. 51-60, 1992

[17]. P.J. Hayes, P.M. Andersen, I.B. Nirenburg, and L.M. Schmandt, "Tcs: A Shell for Content-Based Text Categorization," Proc. Sixth IEEE Conf. Artificial Intelligence Applications (CAIA '90), pp. 320- 326, 1990.