Sentiment analysis in online rating using FP-feed forward artificial neural networks

[1*]Savan Joshi, [2]Anubhav Sharma, [3]Anshul Sarawagi

[1,2,3] CS Department, IES College of Technology, Bhopal, India.

Abstract— In this work the role and importance of social networks as preferred environments for Web mining and sentiment analysis are discussed especially. In this work, selected properties of social networks that are relevant with respect to Web mining are briefly described and outline the general relationships between the two disciplines. The results are outperform and soundly support the main issue of the work, that social networks exhibit properties that make them very suitable for Web mining activities. As a key issue for the successful proliferation on online rating, trust is fast becoming the focus of many research initiatives. This work presents a review and categorization of the trust literature on websites aiming to provide the state of the art as far as research is concerned. Our analysis indicates a lack of research regarding processes for the development of trust and relationship building. The work seeks to fill this gap by proposing a theoretical model for the formation of trust in customer relationships over online rating in websites included-shopping websites.

Keywords— Rule Mining, Classification, Data Mining Algorithms, K-Theory.

I.   INTRODUCTION

In marketing and advertising domains Web Mining is being larger domain. Advertiser needs to analyze performance/ popularity of ads that he/she posted on site. Star rating based mechanism may go fraud, because of robots or automatic responders. So, current system needs to be analyzed using comments & natural language processing. Fraud comments could be removed by using irrelevant comment removal mechanism suggested in work. In this work the role and importance of social networks as preferred environments for Web mining and sentiment analysis are discussed especially.
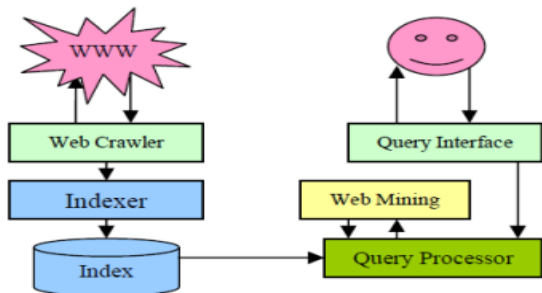

Figure 1: The general architecture of a search engine.

The search engine consists of three main components, known as Crawler, Indexer and Rating [4]. The crawler is also called as a robot or spider that traverses the web and downloads the web pages. The downloaded pages are sent to an indexing module that parses the web pages and builds the index based on the keywords in those pages. An alphabetical index is generally maintained using the keywords.

When a user types a query using keywords on the interface of a search engine, the query processor component match the query keywords with the index and returns the URLs of the pages to the user. But before presenting the pages to the user, a rating algorithm is used by the search engines to present the most relevant pages at the top and less relevant ones at the bottom. It makes the search results navigation easier for the user.

The motive behind this work is to present and analyze the currently important algorithms for rating of web pages such as PageRank and Weighted PageRank and HITS. Second, this work proposes a rating calculation dependent on Topic-Sensitive PageRank and Weighted PageRank. The proposed calculation gives an adaptable way to deal with hunt evaluations utilizing Link examination. For each Web page, the calculation registers a significance score for each subject. At question time, these significance scores are consolidated dependent on the subjects of the inquiry and related setting to shape a composite PageRank score for those pages coordinating the inquiry.

Web mining additionally empower electronic foundations to give better access to different administrations or adverts. At

the point when a foundation makes a promotion for administrations, it provides, however are given by various organizations. Use of mining information will give best to those ways to those entryways. There are three uses for mining in this design.

The main objectives of this work are as follows:

- In this approach, propose a technique which is data mining technique for efficient query processing, it will provide the data containing in it.
- A related keyword and co-extraction is required from dataset.
- Three levels of difficulties are defined for the user existing phase.
- The query needs to get select from the set of form and text values from the dataset.
- The final query will execute by automate system.

II. RELATED WORK

Neha verma et.al. [16] have created an algorithm named the SNEC page rating algorithm which is the semantic and neural based e-commerce page rating algorithm. The site has been rank by utilizing the different fitting highlights to assist the client with picking the best item among numerous items. By this calculation, the organizations can think about their shortcoming and give the client ad libbed item. The outcomes give much precise outcomes however the disadvantage of this calculation is that it incorporates less number of highlights.

Hepp, Martin et. al. [17] has worked for the internet business especially schema.org and Good Relations for analysts and professionals on the trap of the information. In the work, the creator has given a presentation and essential direction on the calculated structure of schema.org. They have made the examples for interest and possession which incorporate the assortment of things and have made a full device chain for creating and devouring the specific information which further utilized for the advanced rating. The creator have likewise talked about the subject like validation (for example with WebID), personality, get to control; information the board issues from the distributer and client point of view and micropayment administrations. The disadvantages of this application is limited to small scale information which isn't appropriate for web based business item evaluating framework.

Sessoms, Matthew, and Kemafor Anyanwu [18] has taken a shot at model and calculations for empowering a Package Query rule on the Semantic Web. The bundle inquiry is the mix of numerous inquiries that gets asset blend on a semantic web. The subclass of such questions is horizon bundle inquiries. Rather than bundle inquiries on a solitary social models, the RDF model have infused the test of deciding the horizon bundle of ternary relations over various joins. The diverse mix of new administrators for horizon bundle inquiries social question administrators and RDF information stockpiling models have built up the four techniques for assessment. The creator needs in the utilization of extra procedures for improvement, for example, prefetching just as the joining of top-k methods.

(Abhijit, 2013) [19] in his paper opined that online business has released one more insurgency, which is changing the manner in which organizations purchase and sell the items and administrations. New approachs have developed. The job of geographic separations in shaping business relations is decreased. With the advancement of 3G and 4G remote correspondence advances, the web economy will keep on developing heartily.

(Raghunath and Panga, 2013) [20] presumed that at first, new web clients would be hesitant to direct any sort of business internet, refering to security reasons as their principle concern. So as to build buyer selection of e-benefits, the wellspring of purchaser perplexity, trepidation and hazard should be distinguished, comprehended and raised. Web based business gives huge open doors in various territories however it requires cautious application for purchaser insurance issues.

(Madhukar Sarode, 2015) [21] in his paper reasoned that web based business is eventual fate of shopping and hole has been decreased among producer and buyer because of web based business. There is immense extension for online business in India however because of feeble digital law, individuals are confronting difficulties in India.

(Saxena, 2015) [22] in her paper inferred that web based business assumes a critical job in Indian culture. It assumes a significant job in redesigning and building up the Indian monetary framework. It offers help to little and medium endeavors to thrive their business. Web based business likewise faces a few difficulties like absence of digital laws and absence of PC instruction and so on.

(H. Ramchandani, 2016) [23] in her paper presumed that the retail web based business deals in 2015 as a level of absolute retail deals in India represented roughly 0.9% of all retail deals in India. Anyway this figure is relied upon to develop in not so distant future and is assessed to achieve 1.4% in 2018. Web based business has extraordinarily affected the matter of physical retailers' particularly little retailers. Global e-retailers are giving a solid challenge to residential ones. Anyway physical retailers still have an excellent remaining in the Indian market as Indian customers by and large prefer to review the products before making any buy.

III. PROBLEM DEFINITION

There are some problems arise in the existing work which are as follows-

1. Large data: this was the major issue when large no. of reviews is available on the shopping websites.

2. Data utilization: because of lack presence of tools which are required to monitor large data was also an issue.

Increasing content: if the no. of websites will increase then the content management will also difficult task. This work focuses on designing a rank reviews by taking some features so that the reviews will be helpful in providing better results and will put on top of the list. In this work the proposed machine learning algorithms that are quite suitable for string data as well as rating. By analysing the portion of reviews that have enough visibility and a high number of votes, the aim to build a model that could predict the helpfulness of reviews with zero or few votes.

IV. PROPOSED WORK

This section presents the Feed Forward Artificial Neural Network (FP-ANN) which calculates the hash label and discovers importance between inputs. The proposed FP-ANN upgrades FP growth calculation with neural networks to sustain the feed forward approach. Key aspects of the proposed FP-ANN are given below.

- Use of ANN Feed Forward Algorithm for the hash label and discovers the importance between the words input.
- In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.
- Generally speaking, if one is given a graph representing a feedforward network, it can always be grouped into layers such that each layer depends only on layers to its left.
- This algorithm creates a hash tag with high rate of exactness and in addition preferable outcome execution over already utilizes ordinary FP-development algorithm for the hash label.
- An info dataset from the given micro blogging stage is taken and processed by different sub process library.

FP-ANN, feed forward layer is based on ANN and it is utilized for calculations to remain productive while finding the information and getting hash label age over the substantial dataset. Figure 2 below gives an idea how the proposed FP-ANN works.
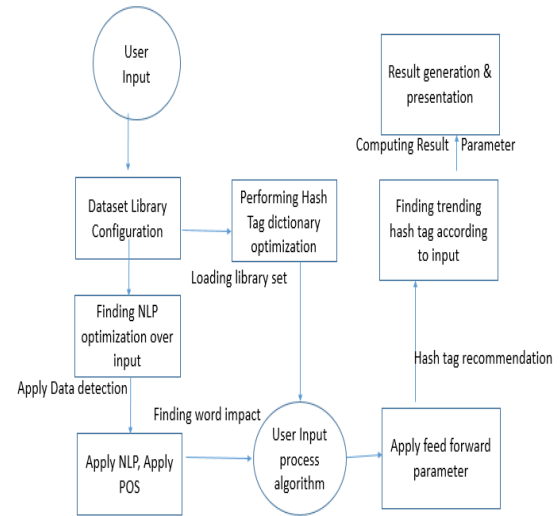

Figure 2: Working of FP-ANN.

Working steps of proposed algorithm:

Step 1: in the initial step the input will be taken from the user as per the need. Listing and loading of all the parameter, component for the simulation purpose and configuration of the entire required scenario framework.

Step 2: in this step the configuration of the library containing datasets will performed also hash tag dictionary optimization will be done in this step. Listing and loading of all the parameter, component for the simulation purpose and configuration of the entire required scenario framework.

Step 3: this step will work on finding the NLP optimization over input.

Step 4: in this step the natural language processing will be applied over the given set of inputs.

Step 5: this is very important step in which the particular algorithm will be chosen to apply on the inputs. Creating an object of all required component. Monitoring de-duplication redundancy and noise verification over the data store and producing the output value of matching. Finding the execution time as per formulae-Execution time = final completion time- initial time; Observing the execution time and thus it effects computational cost for the complete transmission.

Step 6: in this step the feed forward parameters will be applied on the inputs along with the algorithm.

Step 7: in this step the appropriate hash tag will be applied according to the input.

Step 8: computing parameters in this step and then generate proper outputs.

Step 9: End.

The proposed Feed forward artificial neural network (FP-ANN) is profoundly safe and took less calculation time and along these lines computational rate over the organized accessible dataset. ANN in the FP-ANN enables it to perform

quick calculations. It shows its usefulness and calculates similarity measure. These calculations likewise checks for appropriate repetition, usage of more secure and dependable parameters. Pseudo code of FP-ANN is depicted underneath.

V.   EXPERIMENTAL SETUP & RESULT ANALYSIS

All the experiments were performed using an i3-4005U CPU @ 1.70 gigahertz along with 4 Gb RAM on Windows 10. The proposed FP-ANN was implemented in Java using Eclipse IDE with feature selection algorithms. For the experiments a real time twitter data fetching API is used by generating developer key and worked towards real time data fetch storage over local dataset. Further, NLP library configuration is performed using MySQL structured dataset. JAVA:

Computing Parameter: There are mainly two parameter, which is taken for the comparison analysis is taken. Computing parameter such as computation time, computation cost is observed.

Computation Time: Computing time is the time difference which is observed by subtracting final executing time to initial loading time. A time difference between both the times is observed and call as computation time.
Computing time = final execution time – initial time;
Ct=fet-it;

Computation Cost: Computing cost is the total cost which can be observed by monitoring different usage resources and aspects such as bandwidth, data consumption, resources etc.
Computing cost = bandwidth consumption cost +
Resources consumption cost + cost per second;
Cc= bcc + rcc+cps;
Resolving problems:

• The existing algorithm take advantage over previous traditional techniques but still more refinements are required as per todays standard. Thus a better security, hashing mechanism can make it more reliable and executable to tackle with current security and cloud scenario in the world.

• FP growth algorithm makes a repeated computation and accuracy over the repeated value so that it is quite a successful algorithm in generating better outputs.

Computation Time Comparison
Experiments are done to evaluate the proposed FP-ANN and then it is compared to the current state of the art that actually quantifies the performance among its peers. In the experimentations, FP-ANN and current state of the art twitter dataset is used for bringing, marking and plotting of results. Therefore, now it becomes essential to understand total procedure. In underneath tables and figures, gives a review of the performance of both proposed FP-ANN and existing

current state of the art. Experiments are performed for both proposed FP-ANN and current state of art with different dataset size to evaluate computation time (in ms) from 1K tweets to 50K tweets.

Table 1: Computation time comparison

Number of tweets (1000=1K)	Current State of art (Computation time) (ms)	FP-ANN (Computation time) (ms)
1K	14.36	12.2
2K	17.98	16.0
5K	23.12	20.9
10K	43.5	40.89
50K	143.8	132.6

In the above table 1 the comparative study over existing and the proposed is shown.
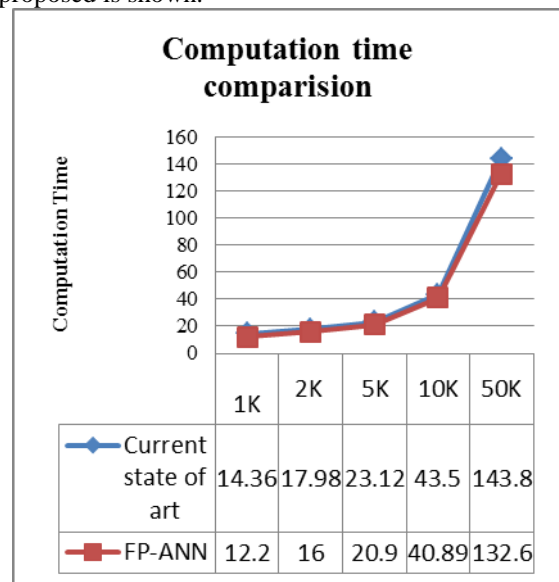

Figure 3: Computation time comparison.

The above table 1 and figure 3 represents computation time for both current state of art and the proposed FP-ANN with variation in number of tweets from 1K to 50K. These results very clearly states that the proposed FP-ANN outperforms the current state of art and reports lower computation time for different number of tweets. For all the values of number of tweets FP-ANN gives better and lower computation time which is very significant in determining the performance of any method. These results also show that the new integrations of feed forward and artificial neural networks are working in tandem to lower computation time.

Computation Cost Comparison
Computation cost is an important parameter to determine the performance of any algorithm. This experiment is performed to calculate the computation cost for processing different sizes of datasets ranging from 1K to 50K for both current state of art and the proposed FP-ANN.

Table 5.2: Computation cost comparison

Number of tweets (1000=1K)	Current State of art (Computation cost)	FP-ANN (Computation cost)
1K	1121	910
2K	2339	1289
5K	2881	2201
10K	3443	3046
50K	8776	6989

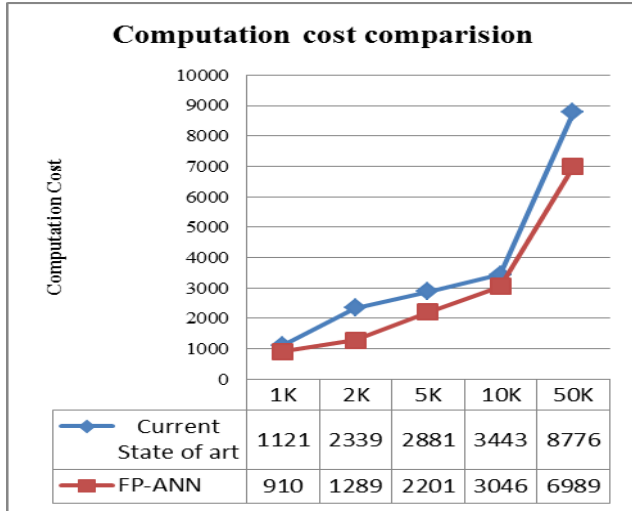In the above table 2 the computation cost comparative study is shown.

Figure 4: Computation cost comparison.

The above table 2 and figure 4 illustrates very clearly put forwards that the proposed FP-ANN performs better as compared to the current state of art and reports lower cost for different number of tweets. For all the values of number of tweets FP-ANN gives better and lower computation cost which is very significant in determining the performance of any method when number of tweets vary from 1K to 50K. Experimental results reflect that mechanism of feed forward with artificial neural network gelled well and lowered computation cost for all the variation in dataset from 1K to 50K.

Overhead Comparison

This experiment is performed with the intent of calculating overhead in the proposed FP-ANN and current state of the art as overhead play detrimental effect in performance.

Table 5.3 Overhead comparison

Number of tweets (1000=1K)	Current State of art (Overhead)	FP-ANN (Overhead)
1K	5454	4340
2K	7845	4908
5K	8081	7668
10K	12897	9880
50K	19912	16569

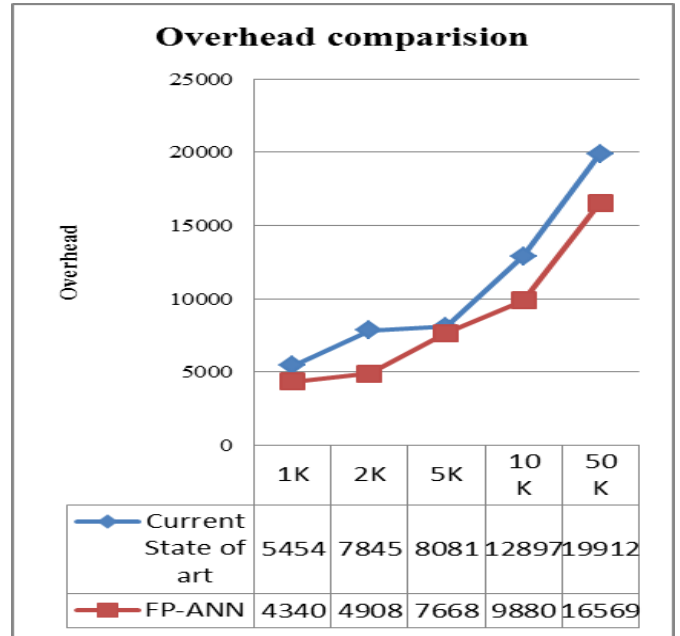In the above table 5.3 the overhead comparison is given.

Figure 5: Overhead comparison.

Experimental results presented in table 3 and figure 5 indicates that proposed FP-ANN reports lower overhead which is better as compared to the existing state of the arts. Trend of result maintain its character and remains same when number of samples in the test set increases from 1K to 5K.

IV. CONCLUSION & FUTURE WORK

Internet has made everyone's life very convenient but has generated humungous amount of data which makes text mining complex and challenging. Various methods have evolved and proved to be helpful in analyzing texts and extracting the information but these suffered from various complexities. This paper presented a Feed forward artificial neural network (FP-ANN) that integrated concepts of feed forward and artificial neural networks and generated different textual patterns from several resources and used hash tag with high rate of exactness. Experimental results very precisely illustrated that FP-ANN outperformed its peers and reported lesser computation time, lower cost and overhead.

As per discussed work and other provided research, following are the future aspects: Integration of more parameter such so to improve the computation time and cost as well as directional parameter can be added. A graphical simulation with different simulation tool can be performing to show effectiveness of proposed approach.

REFERENCES

[1] Alexa, 2014. Retrieved from http://www.Alexa.Com.

[2] BrainSins, (n.d). The State of E-commerce in 2014: Is there any room for new E-commerce models? Retried on 31. 07. 2014 from http://www.brainsins.com/en/blog/stateecommerce-2014-new-ecommerc-models/2643.

[3] Chu, H. A conceptual model of motivations for consumer resale on C2C websites. Service Industries Journal, 33 (15/16), 1527-1543, 2013.

[4] Chu, H. & Liao, S. Exploring Consumer Resale Behavior in C2C Online Auctions: Taxonomy and Influences on Consumer Decisions2007. Academy of Marketing Science Review 11 (3), 1-25, Retrieved on 13.07.2014 from http://www.amsreview.org/articles/chu3- 2007.pdf.

[5] Education Portal, Consumer to Consumer Ecommerce: Definition, Business Model, and Examples, 2014. Retrieved on 12.072014 from http://educationalportal/academy/lesson/consumer-to-consumerc2c-e-commerce-definition-business-modelexamples.html#lesson .

[6] Hannah, B. and Lybecker, K.M. Determinants of recent online purchasing and the percentage of income spent online, International Business Research, 3 (4), 60-71, 2010.

[7] Haque, M.E. Online Shopping usability and customer expectation-A study on online shopping service in India, 2014. Retrieved on 07.07.2014 from bada.hb.se/bitstream/2320/13535/1/2013MAGI03.pdf.

[8] Leonard, L. N. K. Disposition influencers in c2c E-business: purchasing and selling, Journal of Computer Information Systems, 52 (3), 11-17, 2012.

[9] Novicemakers, E-business in India: Status, Potential and Constraints, 2013. Recovered on 06. 07.2014 from Http://novicemarketers.wordpress.com/2013/01/0 6/internet business in-India-status-potentialand-requirements/.

[10] Rami, M., Kleanthi, L., and Nikolaos M. Applying a Multicriteria Satisfaction Analysis Approach Based on User Preferences to Rank Usability Attributes in E-the travel industry Websites Journal of hypothetical and connected electronic business inquire about, 7 (3), 28-48, 2012.

[11] Travis, D. Web based business Usability: Tools and Techniques to Perfect the On-line Experience. Taylor and Francis Group, Oxford, UK. 2003.

[12] Top10 E-business Site Builders, (n.d). An examination of E-trade destinations: A speedy guide, Retrieved on 31.07.2014 from http://www.top10ecommercesitebuilders.com/we b-investigation for-web based business locales a-brisk guideArticle.

[13] Wamuyu, P. K. and Maharaj, M. Components impacting effective utilization of versatile advancements to encourage E-Commerce in little ventures: The instance of Kenya, African Journal of Information Systems, 3 (2), 48-71. 2011.

[14] Wu, F. Li, Hsiao-Hui. Kuo, Yo-Hsin. Notoriety assessment for picking a dependable counterparty in C2C web based business, Electronic Commerce research and Applications, 10(4), 28-36, 2013.

[15] Yu, X., Guo, S., Guo, J., and Huang, X. Rank B2C internet business sites in e-union dependent on AHP and fluffy TOPSIS, Expert Systems with Applications, 38 (4), 3550-3557. 2011.

[16] Verma, Neha, Dheeraj Malhotra, Monica Malhotra, and Jatinder Singh. Internet business Website Rating Using Semantic Web Mining and Neural Computing. Procedia Computer Science 45 (2015): pp. 42-51, ELSEVIER.

[17] Hepp, Martin. The Web of Data for E-Commerce: Schema. organization and GoodRelations for Researchers and Practitioners. In Engineering the Web in the Big Data Era, pp. 723-727. Springer International Publishing, 2015.

[18] Sessoms, Matthew, and Kemafor Anyanwu. Empowering a Package Query Paradigm on the Semantic Web: Model and Algorithms. In Transactions on Large-Scale Dataand Knowledge-Centered Systems XIII, pp. 1-32. Springer Berlin Heidelberg, 2014.

[19] Mitra, Abhijit. (2013). Internet business in India-A Review. Universal Journal of Marketing, Financial Services and Management Research, Vol. 2, 126-131.

[20] Raghunath, Alka. and Panga, Murlidhar. (2013). Issue and Prospects of E-Commerce. Global Journal of Research and Development-A Management Review, Vol. 2, 59-66

[21] Madhukar Sarode, Rajendra. (2015). Fate of E-Commerce in India Challenges and Opportunities. Universal Journal of Applied Research, Vol. I (12), 646-650.

[22]. Saxena, Pallavi. (2015). Internet business and its effect on Indian culture. Universal Journal of Management and Computer Innovations, Vol. 3, 520-522.

[23] H.Ramchandani, Pooja. (2016). Development of E-Commerce in India and its effect on Indian economy. Abhinav International Monthly Referred Journal of Research in Management and Technology, Vol. 5, 33-36.

[24] Goele, S. what's more, Channa, N., Future of E-Commerce in India, International Journal of Computing and Business Research, Proceedings of I-Society 2012 at GKU, Talwandi Sabo Bathinda ,Punjab (Available Websites: http://www.researchmanuscripts.com/isociety2012/7.pdf, 10/11/2012) [25] Hariharaputhiran, S., Challenges and Opportunities of E-Commerce, International Journal of Marketing, Financial, Services & Management Research, Vol.1, No. 3, March 2012, pp- 98-108.