The MindMine Comment Analysis Tool for Collaborative Attitude Solicitation, Analysis, Sense-Making and Visualization

Nicholas C. Romano, Jr., University of Tulsa, Assistant Professor MIS, Nicholas-Romano@Utulsa.Edu
Christina Bauer, Hewlett Packard, Decision Support Systems
Hsinchun Chen, University of Arizona, Director: Artificial Intelligence Lab, Hchen@BPA.Arizona.EDU
&
Jay F. Nunamaker, Jr., University of Arizona, Director: Center for the Management of Information, Nunamaker@CMI.Arizona.EDU

Abstract

This paper describes a study to explore the integration of Group Support Systems (GSS) and Artificial Intelligence (AI) technology to provide solicitation, analytical, visualization and sense-making support for attitudes from large distributed marketing focus groups. The paper describes two experiments and the concomitant evolutionary design and development of an attitude analysis process and the MindMine Comment Analysis Tool.

The analysis process circumvents many of the problems associated with traditional data gathering via closed-ended questionnaires and potentially biased interviews by providing support for online free response evaluative comments. MindMine allows teams of raters to analyze comments from any source, including electronic meetings, discussion groups or surveys, whether they are Web-based or same-place. The analysis results are then displayed as visualizations that enable the team quickly to make sense of attitudes reflected in the comment set, which we believe provide richer information and a more detailed understanding of attitudes.

1. Introduction

The Internet has launched many innovative business activities such as intranets, extranets, virtual corporations, algorithms for customer knowledge discovery, web-spiders, and avatar-inhabited virtual trade shows. These activities have resulted in the ability to gather large volumes of information from potential customers. However, a literature review [1,5,7]. reveals that advances in marketing research have been stalled for at least forty years.

Although computer-based statistical analysis has sped data operations, and bar-code scanning and purchase-pattern analysis allow sophisticated processing of passively collected quantitative data, assessment of consumer attitudes until very recently depended on decades old methods. Problems historically associated with qualitative data gathering techniques such as focus groups, interviews, telephone surveys and mail surveys can create information-discovery bottlenecks that could be alleviated by web-based free response collection and enhanced by subsequent visualization of the results.

Primary (collected firsthand for the study) market research data can be quantitative or qualitative. They usually are gathered through the aforementioned methods, and the accuracy and value of the data still relies heavily on the expertise of information collectors and test designers. The following excerpts indicate how the same types of problems have plagued data collection for decades:

1959 Maketing Research Applications, Procedures, and Cases
…limitations of the survey method are:...Respondents' desire to make a good impression on the interrogator or research group...The make-up of the questionnaire in terms of both content and manner of presentation may lead to bias on the part of the respondent...there is no way of making sure that each question will be asked by each interviewer with the same voice inflection [1].

1983 The Practice of Marketing Research
Limitations of personal interviews include that interviewing expenses in a true random survey can be very high. Interviewer-introduced biases are difficult to detect. Poor interviewers can produce high levels of spoiled returns. Timing of interviews (evenings, weekends) can cause interviewer resentment. Even the most casual study of research theory and practice has indicated that all forms of marketing research work are positively "'hag-ridden' with biases of every type...poor question formulation...poor investigator briefing...careless recording...'hunch proving'" [11].

1993 State of the Art Marketing Research
The (focus) group ordinarily should not have less than seven or more than ten to twelve participants...it is difficult to work with larger groups. At this level, one or two persons tend to dominate the flow of conversation... Limitations of the Telephone Survey...limited interview length...inability to show display materials... The Role of the Interviewer...Accuracy of the data is affected by the skill with which the questions are asked and the finesse with which follow-up and probing questions are handled. Unless the queries are handled in a neutral way, the interviewer's own biases may affect the replies. [2]

1988 Market Trends Research Company
Disadvantages: (mail research) Slowest method of all, taking up to 3 months...No probing of complex questions...Disadvantages (personal interview research) Expensive per interview...Interview bias due to personal appearance or other interviewer effects...Can't reach everyone-out of the way locations are cost prohibitive [7].

These and many other problems that span nearly forty years might be alleviated or even eliminated through use of a Web-based tool that allows free responses from potential customers, demographic sets, or other groups of interest. The high cost of reaching out-of-the way interviewees is eliminated in a Web environment, where the marginal cost per person becomes negligible. Interaction and influence among focus group participants may be improved or enriched, and introduction of bias from interviewers may be greatly reduced or eliminated. For these reasons, using the Web for collecting comments for marketing evaluations could offer substantial cost savings for marketing firms, advertising firms and manufacturers introducing new products.

Once a comment set has been obtained for a stimulus, finding ways to measure attitudes contained therein has proved challenging. Traditionally, marketing firms and new product teams write reports based on overviews of comments or tallies of comment elements, a tedious and labor-intensive process. This project sought to find methods to alleviate data-gathering problems as well to discover new ways to perform comment analysis and display its results.

2. Research Question

Our Overarching research questions is: Can free-response evaluation of multimedia stimuli over the Web produce measurable statistical data regarding preferences and offer content-rich information to the benefit of a marketing team that can then be visualized in a manner that makes sense?

A challenge was to determine whether attitudes contained in a comment set could be visualized and, if so, could they be shown to be quantitatively consistent with or superior to traditional ranking methods.

The first experiment was designed to obtain attitudes on two-dimensional (2D) images from one hundred (100) participants via the Web. The second experiment was designed to elicit the attitudes toward multimedia movie trailers of fifty (50) participants, also via the Web.

3. The Abstract Ideal

The abstract ideal of the application is a completely automated comment analysis tool that uses artificial intelligence to categorize comments and display visualizations in a truly hands-off fashion, such as one developed for the classification of Electronic Meeting Systems output [3,4].

4. Experiment 1: 2D Artwork Images

4.1 Seeking Evaluation Methods & Tools

We designed an evaluation task using GroupSystemsWeb, (GS_{Web}) [8]. The Topic Commenter tool in the application was augmented to support Reference Objects to serve as stimuli for discussion, using 2D art images. Figure 1 shows an example of the GS_{Web} [8] screen with the lunchtime poster.

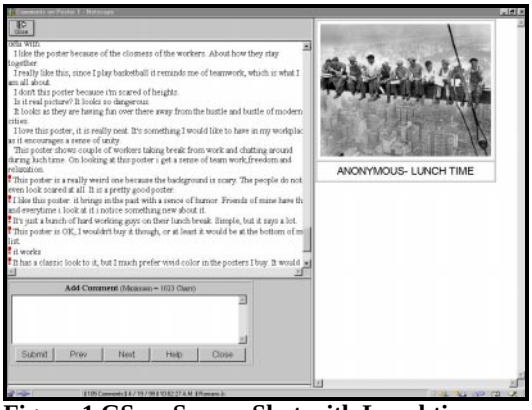

Figure 1 GS_{Web} Screen Shot with Lunchtime

We wanted to find out what a real business considered important in marketing focus groups. With this in mind we performed preliminary requirements gathering from the owners of a wholesale and retail art distributorship. The owners listened to our proposed project and indicated they would like to know which of a selection of popular prints were best liked. As wholesalers of art to other art retailers, they wanted to know which prints to recommend for inventory. They provided us with catalogs, picturing the artwork.

The images selected for comments were:

- "Lunchtime" by Anonymous
- "La Persistenza della Memoria" Dali
- "Omaggio a Grohmann" by Kandinsky
- "Gargoyles" by Parkes.
- "Caffe di Notte" by Van Gogh
- "The Escalator" by Mutter

We then discussed the possible collection of demographic data from the users. Although we had no specific plans for using such data at the initial stages of the project, we surmised there could be possible benefits to marketing firms and new product teams. Therefore, we altered the "business card" feature of GS_{Web} [8], such that fields, such as "fax" and "phone", were replaced with fields more suitable for art evaluation (See Figure 2.)

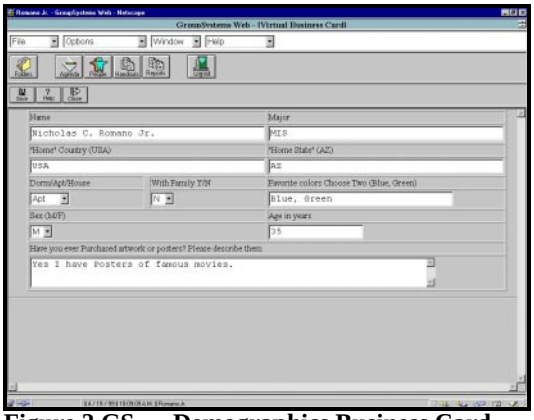

Figure 2 GS_{Web} Demographics Business Card

4.2 Artwork Evaluations

Once GS_{Web} had been suitably altered and the images loaded into the application, we offered students in an Introduction to Computing class an optional for-credit assignment in which they would experiment with GS_{Web} to evaluate art images. Of a class of 142 students, 103 took part in the assignment. Students received an instruction sheet and were informed they could perform the assignment at any time from home, from a university computing lab, or any convenient location.

One of the first instructions was to supply demographic information, which included gender, age, major, living arrangements, and home country. The students had many different majors, including Business, Agriculture, Education, Ecology, Biology, History and Nursing. Females comprised 30% of the participants, and males, 62%. Eight percent did not respond to the gender question. Their ages ranged over 23 years, but 74% were 18 to 24. Although we have not used the demographics for the analysis presented in this paper, we

felt it was important to collect the information for future analyses.

Over a nine-day period, the participants evaluated the images on their own time wherever they chose,. They entered comments via the GS_{Web} Reference Object Topic Commenter card, while viewing each image. At the conclusion of the time period, a reporting feature of GS_{Web} allowed us to print out all the comments, grouped by image number. In addition, approximately two weeks after the test period had expired, participating students were asked to rank the artwork on a 1 to 5 Likert scale, 1 meaning "hate it" and 5 meaning "love it". Zikmund explains that Likert scales ask "*a respondent to indicate the degree of agreement with a statement*" [11]. The information was gathered in order to allow us to compare our method with typical quantitative opinion gathering techniques.

4.3 Attempts to use Arizona Noun Phraser

In the earliest stage, we believed that we would be able to enter the comments into an application called Arizona Noun Phraser [9] for analysis and categorization. However, when the comments were entered into that tool, the resulting categories were not helpful to the task at hand. Although the tool was excellent for topic discovery, it failed to adequately express the attitudes and affect contained within the comment sets.

For instance, the tool left out important comments such as, "it's unique", "it's nostalgic", "it is awesome and reminds me of…", "it symbolized unity", "it is cold and dark". However, the phraser had no problem picking up the key element in the comment, "this poster is for dope heads".

In Figure 3, the Phraser comment output has the following limitations with respect to attitude expression: the word *like* is missing; *black and white photography* is reduced to *white photography*; and *city life* is captured but *reminds me of* is not.

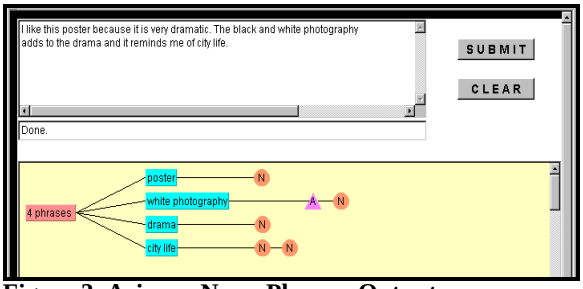

Figure 3. Arizona Noun Phraser Output

In Figure 4 the output is missing several key elements of understanding why a person likes a particular art work: *intriguing, like,* and *you must look closely*.

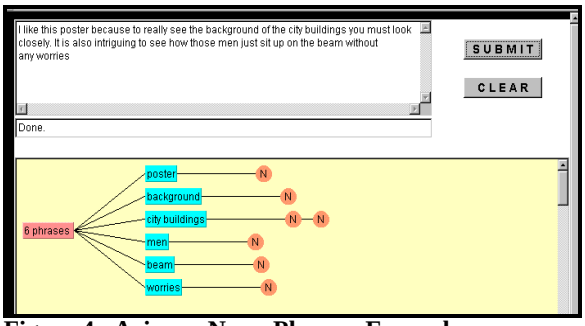

Figure 4. Arizona Noun Phraser Example

How could analysis of the comment set return meaningful information? A quick scan of other available semantic analysis tools revealed the same problems. Although code could be rewritten to recognize adverbs, for instance, we felt that we still wouldn't have information that would represent attitude elements effectively. Nor would such a tool be adaptive enough for the wide variety of goals in attitude research. We thought that visualization would be helpful but realized that it had to be relevant.

It was obvious that we needed to develop a semi-automated method that could later be applied to the creation of an automated analysis tool.

4.4 MindMine Comment Analysis

Once it was determined that Arizona Noun Phraser was not currently a suitable tool to disclose attitudes, we began to develop a new technique for the analysis of the artwork comment set. Approximately sixty of the six hundred comments were scanned and notes were made in a notebook. We broke the comments into parts, and tried to identify a descriptor for each segment: *reminds*, if a student said that the image reminded him of something; *expresses* if they said it showed or made her feel something; *color* if the comment concerned the colors in the artwork. At first, about seventeen categories existed, but upon re-analysis, we reduced that number to twelve (similarities were noticed and terms such as "it shows how..." and "it expresses". were subsequently combined)

Next, we began to use an Excel spreadsheet for the formal analysis. First classified the comments as to directionality based on *like*, *don't like*, or *neutral*. At times, this classification proved so difficult that an additional category called *unknown* was added. Ambiguous or spurious comments such as "*peace*" or "*what the heck is this*" were placed into the *unknown* category.

The next step was to note if the *like/don't like* category had a modifier or intensifier such as "*really like*", "*love it*" or "*hate it*". These terms were put into the *degree* (later changed to *modifier*) column. If the statement was followed by "*it's awesome*", "*it's unique*", "*it's very dull*", then, the terms were added to the *is* column (combined with *modifier* later). Following these steps, the justifications for the *like/don't like* columns were entered. These comments were *because, reminds, expresses, color, however,* (usually an exception to the *like/don't like*), *buy?* (would the viewer buy this), *where* (where would this artwork go or not go), *misc.*(for the unclassifiable) and *for* (suggestions of who might like the artwork).

We used the experience of classifying the comments in an Excel spreadsheet to develop Version 1 of the MindMine Comment Analysis Tool, using Microsoft Access. Access was chosen because of its quick database, forms and report prototyping capabilities. Figure 5 is a screen shot of the application.

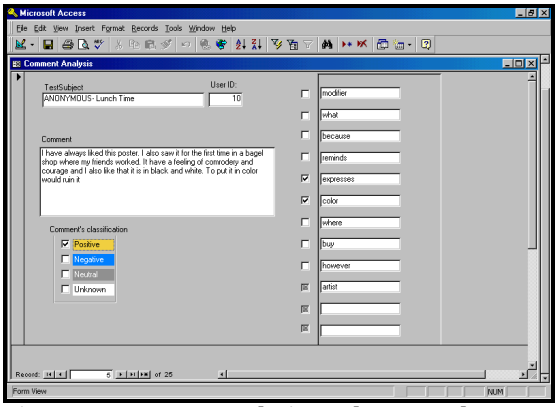

Figure 5 Comment Analysis Tool Screen Shot

When the Access form opens, a selected comment set is in view and a different set can be selected, if desired, using the drop down list box. The user can see how many comments are in each set, offering situation awareness. The entire set, or as many comments as desired, can be scanned or previewed for concept keyword development. The keywords are entered in the blank fields on the right of the screen.

Then, beginning with the first comment, the user chooses a classification from the list as *positive, negative, neutral*, or *unknown*. This action causes the application to insert, respectively, a 0, 1, 2, or 3 into the database field titled *classification*. The screen shows the colors, which are later used in charting the classification. The intent of this feature is let the user become familiar with color associations. Once the classification is established, the user clicks the checkboxes for the keywords that apply to the individual comment. The action of a check inserts a –1 into the database in fields designated as K1-K12, which is correlated with keywords 1 to 12 on the right side of the screen. Mouseover comments can be added to the form to advise the classifier of examples of the keyword's usage.

Once classification and keyword association are completeete for a comment set, the table is transferred to an Excel sheet where the numerical analysis is done. (One version of the Comment Analysis tool, which was developed in Visual J++, automatically produces a chart from the sums of the -1s in the table.) For this project, we chose to enter the data in Excel to ease experimenting with charting variations and to experiment with scoring techniques.

5. Analysis and Development of Quantitative Comparisons

5.1 Analysis of Artwork Comment Sets

Four of the six original comment sets on artwork were selected for use with the Comment Analysis tool. Two were eliminated because of their similarity to other images in the experiment, either in resulting opinions and era of picture or because a subsequent test involved four motion picture trailers.

The four images selected were:
- "Lunchtime" by Anonymous
- "La Persistenza della Memoria" by Dali
- "Omaggio a Grohmann" by Kandinsky
- "Gargoyles" by Parkes.

While using the Comment Analysis tool, users had difficulty deciding between the *because, is,* and *what* categories. Upon subsequent re-evaluation, the category was reduced to *what/why*, which seemed to encompass the three categories. The more specific reasons of *reminds* and *expresses* stayed. This reduction simplified and clarified the keyword association. The new categories and their explanations are as follows:

- **Modifier**- an enhancement to like/don't like, such as "*I love it*", "*it's awesome*", "*I hate it*", "*this is the worst thing I have ever seen.*"
- **What/why**- specific reasons stated for liking or not liking the image such as, "*I like the background*", "*I don't like the dead horse or whatever it is in the foreground.*"
- **Reminds**- the image reminded the viewer of something, such as "*it reminds me of my teammates*", "*it makes me remember that friends are important.*"
- **Expresses**- "*it makes me feel…*", "*it shows how…*", "*I am sad when I see….*"
- **Color**- "*I like the use of black and white*", "*I hate yellow in an office.*"
- **Artist**- the mention of the artist specifically, "*I like Dali.*"
- **Exception**- an opposite statement to the overall feel of the comment, such as "*I really like this picture, but I wish it didn't have so much empty space.*" Exceptions were given a subtractive effect during the numerical analysis.

5.2 Scoring Method

Once all comments had been classified and categorized, we pasted the Access table into an Excel spreadsheet for numerical analysis. The spreadsheet was sorted on the classification column., That is, if the value in that column was 0, 1, 2 or 3, representing *positive, negative, neutral* or *unknown*. Then the number of –1s for each keyword column was totaled as associated with the 0 to 3 value. The keywords associated with negative comments were counted as negative numbers so that the chart would display them below the zero line. This action gave us numerical figures, which were then converted to percentages so that charting across all images could be compared. Next, a comparative scoring method was devised. We created three scores:

- **PNScore**- a score representative of the ratio relationship of positively to negatively related keyword associations, derived from a ratio of keywords associated with positive or negative comment classification. In predominately negative data sets, the ratio is negative to positive and a negative sign is assigned to the number
- **PNRatio**- a score representing the simple relationship of positive comments to negative comments, derived from a ratio of over all positive to negative classification counts. If the set is predominately negative, the score is derived from negative to positive and a negative sign is given to the number. This was not found to be a particularly useful score.

- **P%-** a percentage of positive comments out of all comments in a set (so that the scoring methodology would not be biased).

In creating the scoring mechanism, we planned to ease comparison across images and also hoped that the scoring could later be related to actual sales figures, which had not yet been obtained from The Poster Warehouse,. Figures 6 & 7 show example quantitative results and visualizations.

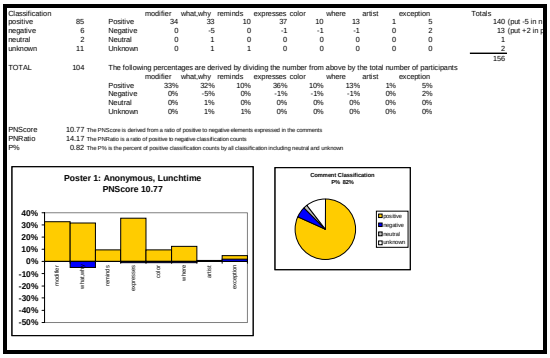

Figure 6 Lunchtime Quantitative Analysis

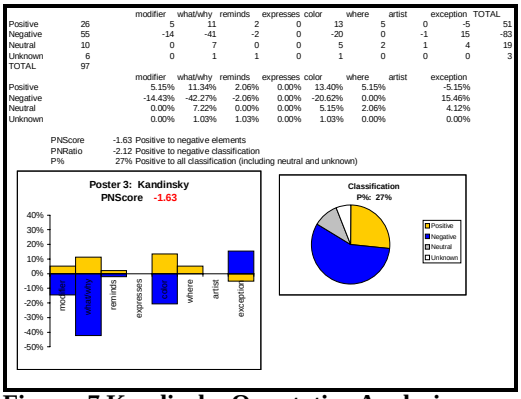

Figure 7 Kandinsky Quantative Analysis

Summary of scoring method

	PN Score	P%
Lunchtime	10.77	82%
Parkes	7.10	54%
Dali	1.08	43%
Kandinsky	-1.63	27%

Inter-rater reliability tests were performed on the data sets using the Comment Analysis tool. On the classification of a comment as positive or negative, which is the primary factor involved in the scoring method, the average reliability rating for all four images was 93%. On the identification of key elements within the comments, the average reliability was 87%.

Comparison of Likert Scale to Mindmine

Student Likert Scale	Mindmine Scoring
1st "Lunchtime"	1st "Lunchtime"
2nd "Gargoyles"	2nd "Gargoyles"
3rd "La Persistenza della Memoria"	3rd "La Persistenza della Memoria"
4th "Omaggio a Grohmann"	4th "Omaggio a Grohmann"

5.3 Sales compared to scoring method

Again, the purpose of this research was not to analyze or discuss the content of the comments as related to marketing or purchasing decisions. Rather, our goal was to present the methodology, scoring, and comparisons to allow professional marketers or new product teams to view our results and determine the method's suitability for their own purposes. However, we feel that it is important to show that the scoring relates to actual sales somehow. Therefore, we obtained the sales ranking for the artwork images analyzed in this phase of the experiment. The sales order is:

5.4 Ranking by Actual Sales

Actual Sales Rank Mindmine Scoring

2nd Gargoyles	1st Lunchtime
3rd La Persistenza della Memoria	2nd Gargoyles
4th Omaggio a Grohmann	3rd La Persistenza della Memoria
1st Lunchtime	4th Omaggio a Grohmann

We revisited our first art analysis using Excel where we used *buy?* as a category (later dropped because of scarcity of related comments).

Would buy Gargoyles	6-8
(2 in doubt include "*can see it in my room*", and "*always wanted a picture like this*")	
Would buy La Persistenza	4
Would buy Omaggio	3
Would buy Lunchtime	1

We discovered that the counts of when viewers specifically stated that they would buy an art work placed "Lunchtime" last, with a count of only one. In fact, the rank order by number of *would buy* comments matched the sales figures. However the number of *would buy* statements ranged only from a low of one to a high of six or eight (depending on how loosely *buy* was defined)., We considered this insufficient support for claiming a definite relationship. However, the fact that only one student stated willingness to buy the artwork indicated that being the group's favorite is significant.

5.5 MindGraphs Creation

Once all the comments had been classified and the keywords associated, we began to experiment with methods to visualize the results. Warm colors were chosen to represent *like* comments, cool colors to represent *don't like* comments, and gray for *neutral* and *unknown*.

This process was repeated for each of the three images. We coined the term "**Mind-Graph**" to convey visualization's, graphic depiction of the group's attitudes along the vertical like-dislike axis and key word associations along the horizontal axis as an alternative to presenting count or percentage-of-preference rankings.

5.6 Discussion and Conclusion

We realized during our study of the comments that viewers can overwhelmingly enjoy a piece of art but not necessarily be moved to buy it. Art is ubiquitous and can be observed in numerous consumer items such as tissue boxes

and umbrellas, not just in relation to the specific purchase of a print to hang on a home wall. This indicated to us that our scoring mechanisms might be best used to determine whether a particular piece of art is likely to be displayed or used in some way rather than simply to predict sales. In fact, we discovered that a New York bagel chain uses the "Lunchtime" artwork , the most popular in our study, in their retail stores. The prediction of sales to individuals might be best obtained by simply asking people if they would buy such a piece of art.

We feel that the experiment demonstrated the rich data resource possible in free responses. For instance, the reasons for the overwhelmingly positive reaction to the black and white poster of men sitting on an I-beam high over New York City were easily observed. Nostalgic elements, perhaps not thought to exist in younger cohorts, were frequently mentioned, as were the feelings of team and friendship. Equally intense and frequent was the expression of dislike for the Kandinsky modern painting. Statements such as, "*my little sister could paint this*" exhibited the hostility that is sometimes observed toward modern art. In addition, some unexpected comments on the abstract Kandinsky, such as "*it reminds me of a mother holding her child*", may not have been anticipated by a marketing firm devising a questionnaire.

Our methodology of scoring the comments using the PNScore and P% mirrored exactly the Likert scale results, demonstrating that the two techniques returned the same rank result. However, our method provided myriad for liking or not liking a work of art, information which can be used by marketing departments and new-product teams to determine what similar or related pieces might be successful or what elements to pursue.

6. Experiment 2: Feature Film Trailers

6.1 Experimental Design

In March of 1999, we designed a new experiment using GSWeb and the MindMine Comment Analysis Tool. We wanted to demonstrate that the tools could also be used for multimedia evaluations. We decided to experiment using movie trailers for films not yet released. and lessons learned from our artwork research

We reconfigured the GS_{Web} frame, which had previously held the artwork images, to open a QuickTime movie player. We then searched www.darkhorizons.com, a web site that links to nearly all available movie trailers, for suitable previews of unreleased and released films. We looked specifically for movies that were unreleased, but had planned release dates in 1999, so that we could compare our scoring results with box office statistics. We settled on the following four trailers, hoping that they would provide enough variety for interesting comparisons:

- **"Muppets from Space"**- The Muppets, in their search for another species like Gonzo, travel to Area 51. There they discover that Gonzo, who has had amnesia since a crash landing in 1947, was the only one of his species to escape from Roswell. Gonzo must decide between returning to his home or staying with the Muppets.
- **"Mystery, Alaska"**- The amateur hockey team from Mystery, Alaska, accepts a challenge to play the New York Rangers. The town pulls together to support the team and becomes the focus of a nationally televised event.
- **"The Mummy"**- An Egyptian priest is mummified alive and cursed for killing a Pharaoh and sleeping with the Pharaoh's mistress. Treasure hunters accidentally resurrect him, unleashing a fury on the world.
- **"Midsummer Night's Dream"**- Shakespeare's love story, complete with fairies, the mischief-maker, Puck, and a donkey's head.

We again altered the Business Card feature of GS_{Web} to include the types of demographics and questions normally found in movie surveys, as obtained from an ACNielson film report. The exception to this was the demographic field "major," which we believed might be of interest to those viewing the statistics. If the entire student population was dominated by Computer Science or MIS majors, the results might be viewed differently than if the students were from the Media Arts program, for instance. The Business Card included the following questions:

- Gender
- Favorite movie type
- How often do you attend movies?
- Age
- Major

Participants again were drawn from an Introduction to Computing class. Instead of written instructions, they were verbally advised to watch the trailers and offer any comments in the space provided. They were told to give their honest opinions, and assured that their answers were anonymous. We felt that downloading the movies to home computers would take too long for most students, so they were given the entire day on two Fridays to view the movies in a multimedia lab on the campus at the University of Arizona. Figure 8 shows the GS_{Web} Topic Commenter screen with a movie trailer running as the reference object.

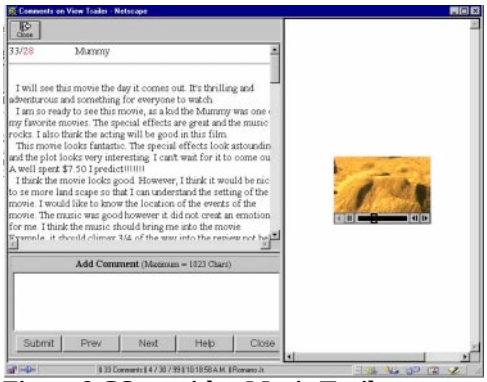

Figure 8 GS_{Web} with a Movie Trailer

We then followed the same procedure that had been developed during the artwork evaluation. We pasted the comment sets in the Comment Analysis Tool Access database. We scanned the comments from the report created by GS_{Web} for concepts and trends. During this period, we developed the idea that movie trailers and artwork cannot be rated in the same manner. As mentioned earlier, art can be appreciated and enjoyed and can enhance sales of other items, without the specific purchase of a print to hang on a home wall. Because of this, positive/negative reactions to an image are critical. Movie trailers, however, have no other

purpose than to entice the viewer to see the film. Therefore, instead of making the positive/negative reaction the key to the charting and quantitative analysis portions of the experiment, we replaced *will see/won't see* with the following classification categories:

- Will see
- Won't see
- Wait for video
- Unknown

We theorized that it did not matter if the viewer had a positive reaction to the movie clip if it did not make him/her want to see it. This effect was noticed when viewing the comments about "Muppets from Space." Although the response to this trailer was overwhelmingly positive, few students said they would see it, whereas many specifically stated that they couldn't wait to see "The Mummy." Since movie attendance is, in fact, the goal of the movie trailer, we felt that we had justified our method change. We also believed it to be significant if a viewer declared that he or she wanted to see the movie without being asked.

The process of identifying categories or keywords resulted in the following selection:

- **emotion**- "this is a feel-good movie", "I always like to root for the underdog"
- **actor/character**- used when a specific actor or character is mentioned, for instance "I'll see anything with Brandon Frasier", or "Miss Piggy rules!"
- **director/writer**- specific mention of the writer or director, such as "I've read many of Shakespeare's plays and I want to see this"
- **music**- specific mention of the music
- **special effects**- specific mention of the special effects
- **reminiscent**- "this reminds me of…(another film)"
- **with someone**- "I would take my little cousin to see this"
- **the genre**- "I like sports movies"
- **scenery**- "the mountains and Alaskan scenery look great"
- **plot/subject**- "I like Egyptian history so I think I'll like this film"

The Comment Analysis tool was used to classify and assign keywords to each comment in the four sets. Again, each resulting Access table was copied and pasted into an Excel spreadsheet for numerical analysis and charting.

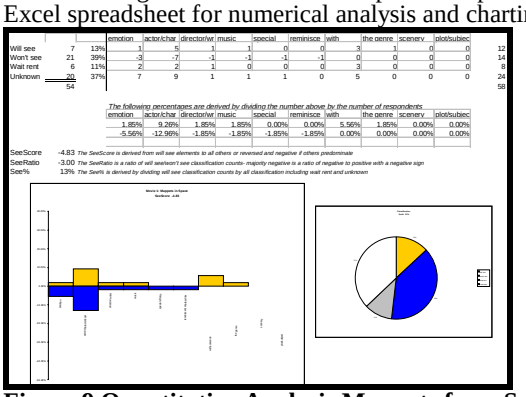

Figure 9 Quantitative Analysis Muppets from Space

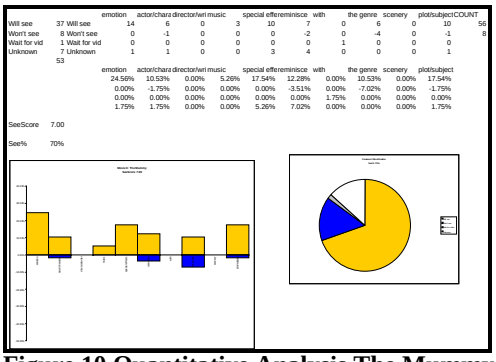

Figure 10 Quantitative Analysis The Mummy

6.2 Description of scoring method

The scoring used for the movies was similar to that used for the artwork experiment. But, instead of using the *positive/negative* classification upon which to sort the keyword elements, we used will *see/won't see.*

- **SeeScore** was derived by dividing the count of elements associated with *will see* by the count of the elements associated with *won't see*, *wait for video*, and *unknown*.
- **See%** was derived by taking a simple percentage of the whole set of comments indicating plan-to-see.

Inter-rater reliability tests were conducted on the data sets using the Comment Analysis tool. The reliability rating for the factors involved in the will see/won't see calculations were 88 per cent. However, the association with specific keywords dropped to 73 per cent. We realized that users struggled with "emotion." One rater found many emotional elements; others did not. We decided that emotion was too vague. Also the difference between genre and plot/subject proved confusing. We feel that further studies of the comments would produce a better keyword set, which could be reused for movie evaluations.

6.3 Comparison of scoring to per screen statistics

At the time of this writing, only three of the four movies had been released, "The Mummy", "Midsummer Night's Dream" and "Muppets from Space." The box office totals from only the first week were compared because we believed word-of-mouth and other advertising would thereafter have more of an effect on attendance than the movie trailer. The first week box office results were:

- The Mummy
  - Gross $43,369,635
  - Per screen average $13,515
- Midsummer Night's Dream
  - Gross $4,285,620
  - Per screen average $ 3,968
- Muppets from Space
  - Gross $6,686,522
  - Per screen average $2,131

"The Mummy" out-grossed "Midsummer Night's Dream" by a factor of 10. "The Mummy" also had higher per screen averages by a factor of 3.4. To compare, we have listed the scores and box office factors side by side.

Movie	SeeScore	See%
Mummy	7.00	70%
Midsummer	-2.36	38%
Muppets	-4.83	13%
Mystery	3.20	57%

We believe that the fact that our population consisted of college students increased the attendance score of for "Midsummer Nights Dream" over that that might be found in the general movie-going population. Similarly, participants may have underrated "Muppets from Space" because of their general lack of offspring. In comparing "Midsummer Night's Dream" and "The Mummy," two films released about the same time, SeeScore and See% ranked the same, as did their per screen factor comparisons and their total box office gross. The SeeScores had a 9.36 to 1 factor and the Box office statistics a 10.1 to 1 factor, repectively. The important point is that the methodology described in this paper accurately reflected the greater popularity of "The Mummy" over "Midsummer Night's Dream" in a per screen and box office relationship.

6.4 Comparison of Scoring to Likert Scale

We collected a ranking from the students on a separate paper following the GS_{Web} test in order to compare the two methodologies. The rank sheet had each movie listed with the numbers 1 to 5 underneath the movie title, as follows:

The Mummy
1 2 3 4 5
Hate it…….Love it

We then changed the 1 to 5 to –2 to +2 during analysis in order to obtain a comparative score and in order to chart the reaction and appropriately represent the negative feeling. The average of the results were:

The Mummy	+0.67
Midsummer Night's Dream	+0.33
Mystery, Alaska	+0.17
Muppets from Space	-0.41

6.5 Movie Trailer Experiment Conclusion

We believe that after the release of *Mystery, Alaska*, our method of scoring, called the SeeScore, will prove superior to the Likert scales. We expect *Mystery, Alaska* to outperform *Midsummer Night's Dream* by a large dollar amount in the box office statistics. We do not have a specific explanation for the failure of the Likert scaling to predict the correct order of popularity of the movies. However, we feel that it may be due to the richer information contained in the free responses. The viewers, themselves, may not be able adequately to articulate their views on a 1 to 5 scale.

7. Future Directions

7.1 Future human factors considerations

As one of the most important implications of this project is the possibility of developing automated methods for the techniques described, discussion of future directions is centered first on further development of the Comment Analysis tool.

7.2 Design Directions - Single-User Tool

The Comment Analysis tool has many aspects for further development and enhancement.

1. **Statistical, demographic**: classification and keyword cross-analysis: This feature would allow cross-analysis between users, their demographic data, their comment classifications and even individual keywords.

2. **Dynamic thesaurus:** The dynamic thesaurus works when an analyst enters a new keyword. When the system detects the new word, a star burst view of similar words appears. The user selects which words should be included with the entered keyword. For instance, the first time that a user determines that the word *reminds* is applicable to a comment set, a star burst of words similar to reminds would appear in a list form That list may initially contain words such as *remember, recall, recollect*. At this point, the user could add other words he or she determined to be associated, such as reminisce. This set becomes a customized, user-created thesaurus, which could be applied only to the current set under analysis, or saved for future use. The system could then perform a search of all the comments for the varying forms of the selected root words. This would speed the keyword assignment process.

3. **Phrase Binning :**Users might find a phrase binning feature helpful. The analyst would be allowed to highlight a phrase from a comment, and place the phrase into a virtual box which is assigned to a keyword. In this manner, the analyst could later view all of the phrases that were associated with "where" in the artwork viewing test, for instance. The user could then see at a glance test takers' ideas about the best place for the picture to be placed, such as "in an office", or "in my den" without reading the entire comment again.

4. **QuickViz - Intermediate visualizations :** An analyst might find intermediate visualizations an aid to the classification process. Using this feature, the analyst might notice that the visualization has stabilized, making further comment-classification is unnecessary. This might be a help in very large data sets. As mentioned previously, caution is required for potential bias introduction.

5. **Demographic analysis:** As demographic data has been collected on the participants in both experiments, cross analysis and graphing is certainly possible and could offer insights to the analyst.

6. **Group Participation:** One future direction might be to facilitate discussion that enables the focus group to become aware of the opinions of others and thus provides additional information through idea triggering and other group process gains.

7.3 Cautions for an Automated Version

Careful consideration of the decision-making processes of potential users of the system, might lead to a conclusion that a semi-automated device is preferable to a fully automated system for several reasons, but the following possibilities should not be ignored.

- Biases and heuristics of the creator of a fully automated keyword association tool might influence the criteria for classification and prevent the user's creating a tailored keyword set designed for a specific goal

- Users may need to develop their own industry- or project-specific keyword sets to establish heuristics best suited to their own environment. However, such keyword sets should be reviewed or approved by management to avoid individual bias from entering into the test results at this stage (at least any will be corporate!).
- A more fully automated device could reduce workload, however, by using suggested keywords obtained from text scans. A requirement for final human approval of keyword assignment could be an on/off feature, providing an adaptive system that could be varied based on workload. Such a version would introduce further considerations such as:
- The corporation's acceptability heuristic,
- Whether the level of automation will be mandated by the company,
- Acceptance of the accuracy of the automated results - -consistently inaccurate keyword associations will result in non-use

Finally, the decision aid essentially must enable a combination of the human decision maker and the automated aid team to act together as a high performing team whose capabilities exceed those of either element alone.

7.4Future directions of the methodology

Domains: Developing reliable scoring methods for various industrial applications will require working in each specific area to obtain the optimal set of concepts and other criteria on which to base quantitative measures. Qualitatively rich-response data sets are complex and multifaceted, and may require an understanding of the domain in which the discussions are held.

Weights: While developing scoring for response data, some concepts, keywords or factors may require assignment of weights that reflect their relative importance, determined by such factors as relationship to predicted sales, importance to information discovery based on specific goals, and other bases.

Unknowns: Comments classified as unknown (either positive/negative or will see/won't see) are a direct result of not requiring the viewer to be specific. We theorize that undirected responses are richer because they free the viewer from restricted thought. However, with more experience, we may find that requiring a direct indication of whether or not the viewer plans to see a movie (a "required" field perhaps), may be more important than release from presumed mental restrictions. It would be interesting n to compare the responses from a group who have been instructed to state their plans to see a with those movie within their comment given no such requirement to see which provides less information. l

8. Revisiting the Research Questions

The following research questions were posed at the beginning of this project.

Q. Can multimedia free-response evaluation comments be successfully collected over the web?

A. We concluded that GS_{Web} was used successfully to gather free-response comments from users who performed their evaluations from home and public labs, as well as in a controlled lab environment. We also showed that both 2D and multimedia objects could be evaluated over the Web.

Q. Can free responses be evaluated to produce measurable and comparable data for analysis?

A. This project demonstrated a set of techniques to give quantitative measures to qualitative data. We also showed that the measures of the movie trailers could be matched to real box office statistics.

Q. Do the benefits to free-response collection, besides the economic and collection benefits mentioned in the introduction, also include content-rich information that is beneficial to marketing and advertising professionals, and new product teams?

A. We believe this project has shown that rich content can be obtained from letting evaluators freely offer their opinions without the structured guidance of traditional question/ranking surveys. We also believe that our method will prove superior to Likert scaling for prediction of movie success from trailers. It is now up to professional marketers and advertisers to determine the specific value of these results.

Q. Can the attitudes contained in a comment set be visualized?

A. Yes.

References

1. Alevizos, J. P. *Marketing Research Applications, Procedures, and Cases.* Prentice Hall, Inc., 1959.

2. Blankenship, A. B. B., G. E. *State of the Art Marketing Research.* NTC Business Books, 1992.

3. Chen, H., Houston, A., Nunamaker, J.F., Jr., and Yen, J. Toward intelligent meeting agents. *IEEE Computer*, 29, 8, (1996), 62-70.

4. Chen, H., Titkova, O., Orwig, R., & Nunamaker, J. F. Jr. Information visualization for collaborative computing. *IEEE Computer*, 31, 8 (August), (1998), 75-82.

5. Davies, A. H. *The Practice of Marketing Research*. London: Heinemann, 1983.

6. Deese, J. Conceptual Categories in the Study of Content, G. Gerbner, O. R. H., K. Krippendorff, W. Paisley, P. J. Stone Eds. (ed.), *The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques.* New York: John Wiley & Sons, Inc., 1969.

7. Market ,Trends Research Company Available at http://www.markettrends.com/p0000049.htm, 1998.

9. Romano, N. C. Jr., Nunamaker, J. F. Jr., Briggs, R. O., and Vogel, D. R. Architecture, design, and development of an HTML/Javascript web-based group support system. *Journal of the American Society for Information Science* 49 (7), pps. 649-667, 1998.

10. Tolle, K. M., & Chen, H. Comparing Noun Phrasing Techniques for Use with Medical Digital Library Tools, (In Press.)

11 Zikmund, W. G. & d'Amico, M.. *Marketing*. New York, NY: West Publishing Company, 1996.