Volume 5, Numbers 1 and 2 **ISSN 1554-5393**

JOURNAL OF STRATEGIC E-COMMERCE

An official Journal of the
Allied Academies, Inc.

David Wyld
Co-Editor
Southeastern Louisiana University

Randall Settoon
Co-Editor
Southeastern Louisiana University

Academy Information
is published on the Allied Academies web page
www.alliedacademies.org

The Allied Academies, Inc. is a non-profit corporation chartered under the laws of North Carolina in the United States. The Academy is an association of scholars and practitioners whose purpose is to advance the knowledge, understanding, and teaching of e-commerce and e-government throughout the world.

Whitney Press, Inc.

Printed by Whitney Press, Inc.
PO Box 1064, Cullowhee, NC 28723
www.whitneypress.com

The *Journal of Strategic E-Commerce* is owned and published by the Entrepreneurship Group, LLC, PO Box 2689, Cullowhee, NC 28723, USA, (828) 293-9151, FAX (828) 293-9407. Those interested in subscribing to the *Journal*, advertising in the *Journal*, or otherwise communicating with the *Journal*, should contact the Allied Academies Executive Director at info@alliedacademies.org.

JOURNAL OF STRATEGIC E-COMMERCE

Volume 5, Numbers 1 and 2, 2007

JOURNAL OF STRATEGIC E-COMMERCE

CONTENTS OF VOLUME 5, NUMBER 1

JOURNAL OF STRATEGIC E-COMMERCE

CONTENTS OF VOLUME 5, NUMBER 2

LETTER FROM THE EDITORS

We are extremely pleased to present this issue of the *Journal of Strategic E-Commerce*, an official publication of the Allied Academies, dedicated to the advancement of knowledge, understanding and teaching of e-commerce throughout the world. The editorial mission of this journal is to publish empirical and theoretical manuscripts which advance the e-commerce initiatives.

The Allied Academies, Inc., is a non profit association of scholars whose purpose is to encourage and support the world-wide advancement and exchange of knowledge, understanding and teaching. The *JSEC* is a principal vehicle for achieving the objectives of the organization.

As has been the case with the previous issues of the *JSEC*, the articles contained in this volume have been double blind refereed. The acceptance rate for manuscripts in this issue, 25%, conforms to our editorial policies.

The Editors work to foster a supportive, mentoring effort on the part of the referees which will result in encouraging and supporting writers. They will continue to welcome different viewpoints because in differences we find learning; in differences we develop understanding; in differences we gain knowledge and in differences we develop the discipline into a more comprehensive, less esoteric, and dynamic metier.

Information about the Allied Academies, the *JSEC*, and the other journals published by the Academy, as well as calls for conferences, are published on our web site, www.alliedacademies.org. In addition, we keep the web site updated with the latest activities of the organization. Please visit our site and know that we welcome hearing from you at any time.

David Wyld
Randall Settoon
Southeastern Louisiana University

ARTICLES for Volume 5, Number 1

ONLINE RECRUITMENT: ATTITUDES AND BEHAVIORS OF JOB SEEKERS

Patricia C. Borstorff, Jacksonville State University
Michael B. Marker, Jacksonville State University
Doris S. Bennett, Jacksonville State University

ABSTRACT

The Internet has changed the way companies conduct business, including how they attract and recruit employees. As a result, online recruiting has become a major Internet business tool. Companies can recruit online with their own websites, job boards or resume banks, newspaper classified ads, and specialized job boards or professional certification sites. Effective online recruiting processes increase firms' competitive advantage through increased efficiency and lower costs, and offers benefits and opportunities to job seekers. This research investigates the perceptions and behaviors of job seekers concerning the use of the Internet as a recruiting source. We found citizens to be more comfortable with online recruitment and used it more frequently than did non-citizens. Older applicants and those with work experience also applied more often and made more job searches. Minorities applied more frequently for job online than did whites. We found no differences in gender and online recruitment behavior or attitudes.

INTRODUCTION

In the current volatile business market, one of the most distinctive competitive advantages companies can gain and sustain is their human resources. Thomas and Ray (2000) believe that the single most important determinant of organizational effectiveness is the ability to attract, hire, and develop capable talent. To be able to compete, firms must be able to find and retain the best available employees. This is difficult due to the shrinking availability of qualified labor. Furthermore, the rise of computer technology and the Internet has changed the way businesses compete. One area that has been changed drastically by information technology is e-recruiting or online recruiting.

Online recruiting is the process of recruiting through commercial job sites or company websites that promotes employment opportunities and retrieves potential employee information (Lin & Stasinskaya, 2002). Online hiring is used to post jobs, accept resumes, administer screening tests, and correspond with job applicants (Flynn, 2002). In the past, most organizations used employee referrals, newspaper ads, and traditional employment agencies to advertise job vacancies. Today, the Internet is a popular method to recruit potential employees, with the competition for qualified talent being online. Over 90% of *Fortune 500* companies use some form of online recruiting (Feldman & Klaas, 2002). One survey showed 85 percent of companies with 500 or more employees in North America have an online recruiting program (Schweyer, 2003). Job seekers are also conducting their searches online. Over 52 million Americans have used online job searches, with over 4 million doing so daily (Jansen, Jansen, and Spink, 2005). Also, studies based on hiring practices reveal that 51 percent of new hires in 2005 were generated from internet sources (Hamilton, 2006). The job database Monster.com claims to have more than 48 million resumes in their database, with almost 4 million people visiting their site on Mondays alone (Backhaus, 2004). Similarly, Hadass (2004) found that the population of users of this website has increased to 6.5 million and is constantly growing. These numbers do not include the 40,000 other job boards available on the Internet.

This increasing use of online recruiting and hiring as a business tool has not only changed the way companies recruit employees and how job seekers search for jobs, it has also impacted both parties involved. This research reviews current online recruiting and hiring practices and examines the attitudes and behaviors of online job seekers in their use of company websites and the online job boards.

METHODS AND PRACTICES EMPLOYED IN ONLINE RECRUITING

Traditionally, organizations depended on low-tech, time-consuming recruiting methods such as newspaper ads, employee referrals, and employment agencies to locate and attract qualified candidates. While these methods may have served employers well in the past, the need to be dynamic and progressive is predicated on the global and competitive business environment of the 21st century. Today, companies view the Internet as a way to extend their business at little cost by offering information and advertisement as well as conducting business on their websites. According to a National Online Recruitment Survey (NORAS), seventy-two percent of organizations acquired their resumes or curriculum vitas through various forms of electronic communication methods available through the internet (Zall, 2000).

Online recruiting includes at least four identified techniques, including company websites, general online job boards (sometimes referred to as internet job search engines, resume databases or resume banks), online newspaper classified advertisements, and job postings through specialized job boards or professional certification cites.

The two most popular tools available for online recruiting are company websites and Internet job boards. The most popular online recruiting tool, the company website, is the most convenient of all. Simple HTML editing software allows HR staff members to take *Word* documents and post job requirement information directly onto the website.

These websites help companies perform the first two major functions of the hiring process; attracting applicants to the company through the posting of available jobs, and receiving applications and resumes from the internet. Company websites that were originally used as public relations and commerce tools are now creating a virtual labor marketplace (Wyld, 2005). In a survey by iLogos.com, 96 percent of the recruiters polled reported posting jobs on their company's websites. The Value Creation through Corporate Careers Websites study from iLogos Research found that 92 percent of *Fortune 500* companies have a website solely for careers (Co.-Specific Career Web Sites, 2003). The results of several surveys and studies suggest that company career websites are very attractive. A survey of job seekers conducted by CareerXroads found that 92.4 percent of respondents were likely or very likely to visit a company's website to find out more about the organization. The study noted that although 85 percent of the respondents had gone to company websites for other reasons, they found themselves checking out available jobs (Agnvall, 2005). Another survey of 70 leading US companies performed by DirectEmployers Association showed that nearly 25 percent of survey respondents' new hires in 2005 were found using corporate career websites (Minton-Eversole, 2006).

The second popular online recruiting method is Internet job boards. Job boards are websites that act as places for companies to look for job seekers and job seekers to look for jobs. These online job boards function much like an electronic version of the traditional newspaper ads that companies have been using for years (Volpe and Tucker, 2004). Companies can post job ads on these sites in addition to using the sites' resume banks to search for potential applicants they would like to contact. On the other side of the search, potential applicants are able to access the job board to view job postings and submit resumes and applications to the companies they would like to work for. Today, there are many job boards excelling in online hiring, such as Dice, Monster, Hot Jobs, CareerBuilder, and Vault (McLean 2006). The online job board Monster.com reports that for April, 2006, a total of 28 million visitors spent an average of 12 minutes on the site with over 12 million visitors described as "quality" (Monster Advantage, 2006). The Monster Fact Sheet (2006) reported having 61 million job seeker members worldwide, a resume database of more that 52 million resumes, and over 200,000 member companies. While job boards have focused on recruiting management and white-collar workers, recent television ads by Monster.com have begun promoting the site as an option for the blue-collar job seeker. Both online methods, company websites and job boards, are important tools for organizations and job seekers wishing to become active in the online job market and the 52 million resumes Monster.com has stored on its site and the 40,000 other resume-posting websites points out that companies cannot ignore online job boards as an essential recruiting tool.

IMPACT OF ONLINE RECRUITING ON BUSINESSES AND JOB SEEKERS

The use of online recruiting impacts significantly businesses and job seekers who utilize this method in the hiring and job search process. Businesses are finding many favorable reasons to use online recruiting including increasing efficiency, lowering recruitment and hiring costs, attracting more qualified applicants and simplifying the entire selection process. Using online recruiting, a company is able to streamline and increase the efficiency of its HR functions, leading to a decrease in the cost of recruiting and hiring new employees (Karakanian, 2000).

Regarding increasing efficiency, the move to online recruitment provides managers with rapid access to the information essential when planning, directing, and addressing staffing needs. State of the art resume databases provide recruiters and human resource managers easy access to the best talent in the workforce. Employers can tailor questions and conduct searches for candidates based on specific skills, knowledge and abilities required for the position. From the database of information, they can extract a list of individuals that best meet the needs of the company. In addition, providing facts about organization culture, environment, and practices can increase the visibility of the employer and enhance the chances of properly aligning the employer with a suitable employee. (Zall, 2000) Furthermore, company websites can also be used to receive electronic resumes and online applications. The collection of resumes and applications online allows companies to quickly acquire large amounts of data on potential job applicants (Karakanian, 2000). The use of job agents allows HR departments to target and identify quality candidates with more efficiency. Utilizing Internet prescreening tools and completing and submitting the job application online streamlines recruiting and talent deployment processes, thereby cutting the length of the hiring cycle (Co.-Specific Career Web Sites, 2003). This data collection and the control a company has over its procedures creates a process that allows a company to collect the data it needs at a quicker pace than with traditional methods. Many different applications available allow a company to evaluate the information posted prior to its movement to the HR department.

Recruitment activities managed in an electronic format can lower recruitment cost by prescreening applicants, processing applications, and maintaining applicant information for forthcoming employment opportunities (Zall, 2000). Hogler, Henle & Bemus (1998) reported that it costs a company nearly one-third of an employee's salary to replace that employee. The Internet assists companies in potentially decreasing that cost by an estimated 20 to 30 percent (Menagh, 1999). Lee (2005) reported that Dow Chemical was able to cut costs per hire by 26 percent through utilizing a careers website. By eliminating the need for applicants to be at a physical location, online testing allows for companies to locate qualified applicants from a wide range of geographic locations prior to costly onsite visits (Mooney, 2002). This means that companies who engage in online hiring gain a distinct advantage over their competitors who are not online.

Additionally, online recruiting exposes the recruiting company to many more applicants searching for jobs. With nearly 200 million people using the Internet (Gale, 2001), it is very likely that some of that traffic will be crossing a company's website. The company website is an invaluable tool in the recruitment of the active job seeker. A major step for the current job seeker is analyzing the company online. Providing facts about organization culture, environment, and practices can increase the visibility of the employer and enhance the chances of properly aligning the employer with a suitable employee (Brice & Waung, 2002). Cappelli (2001) noted the GE website that linked applicants to information regarding what working for their company would be like. Companies are also using their websites to link potential job applicants to various other sites that could attract applicants, such as community sites and local attractions (Wyld, 2005). For the

Internet users that are simply surfing the Internet and not actively looking for a job, the Internet is a great place to sell a company's image and culture. The website is a virtual brochure that can be updated regularly to attract any site visitor to apply for a job (Greengard, 1995). Whether the surfers are on a web page to look for a job, buy a product or just find more information on a company, it is an opportunity that cannot be missed in recruiting talent.

Job seekers submitting, and companies accepting resumes online eliminates the need to fax or mail resumes and makes the process of submitting a resume instantaneous (Wyld, 2005). Companies are allowed to receive thousands of resumes in short periods of time, which expedites the process of filling vacancies. When a resume is submitted in an electronic form, it allows a firm to quickly evaluate large numbers of resumes in a limited number of hours through key-word searching. Using a key-word search reduces large numbers of resumes received online to a small number of qualified candidates (Mohamed, Orife, & Wibowo, 2001). Once these candidates have been identified, the electronic resume can easily be transferred to the relevant departments involved in the hiring process (McCune, 1998). The direct submission and review of resumes lowers the cost of receiving and evaluating potential job applicants.

After resumes have been evaluated and a pool of qualified candidates has been established, companies can test and screen their applicants online (McCune, 1998). Twelve percent of companies recruiting online administer tests online (Cappelli, 2001).

Software programs allow a company to quiz applicants electronically and determine the suitability of an applicant to adequately perform the job (McCune, 1998; Mooney, 2002).

Personality traits and integrity can be analyzed online to evaluate the fit of an applicant to a position prior to an interview (McCourt-Mooney, 2000a). The HR department also uses the Internet to search for and validate such information as criminal records, references, and prior employment.

IMPACT ON BUSINESSES

Although businesses receive many benefits from recruiting online, there are challenges, including loss of potential candidates from application overload or technical problems, loss of valued employees, loss of "personal touch" by recruiters with potential applicants, and in some cases, increased costs in the recruiting process.

Although employers are able to provide more job listings and recruit more applicants faster using the internet, many have found it difficult to satisfy their job recruitment needs. Quantity does not always turn into quality. With millions of resumes circulating around the Internet, and with the ease of applying for a job online, employers are receiving applicants with below standard skill sets, causing unnecessary work for the hiring manager who has to sift through and eliminate non-qualified candidates. Also with the proliferation of online job postings, it can be difficult for the employee and the employer to find each other. This is especially true with online job boards such as Monster.com and CareerBuilder.com. A simple internet search of the word “job” revealed that there were over two billon web pages with this word (www.google.com). These sites offer a wide variety of jobs and are useful for employers wishing to reach a wide audience. However, these large sites can become a quagmire of too many people and can quickly swamp a HR Department with too many applicants.

Another challenge for online recruiting lies within the technical abilities of the company websites to provide an effective database that is functional, user friendly, and allows easy data entry for users as well as simple data retrieval by human resource professionals Because the high volume of career sites on the Internet creates natural competition, different sites must use various strategies for attracting visitors to submit their resumes, including different styles of organizational layout, color scheme and navigation (Lin & Stasinskaya, 2002). Also, career sites must be designed with advanced capabilities, use of artificial intelligence, and keyword matching systems that will accurately reflect search criteria. These attributes are proving to be difficult to accomplish by some company sites.

The speed with which technology is changing presents another challenge with online recruiting. On average, new and faster computer technology appears every six months, causing difficulty for the employer to stay literate on the latest computer software, hardware, and Internet technology. This constant change also requires the HR staff to become more technology savvy or more dependent on the company’s IT department.

Another problem associated with businesses using online recruiting is a decreasing trend of personal attention. A survey by Partners (2000) identifies personal interaction as very important to both employers and job seekers. For employers, personal interviews serve as good indicator of an applicant’s composure, allowing the employer to gauge a potential employee’s ability to work under pressure by observing his/her composure during an interview. This study shows that many job seekers consider online resume posting inefficient due to, among other reasons, lack of personal attention. Today, many HR professionals find themselves viewing people as 1's and 0's on computer screens, drastically altering the traditional hiring process and possibly the attitudes and behavior of job seekers and potential employees (Agnvall, 2005).

The use of online recruiting may cause companies to lose valued employees. Search engines allow companies to view and evaluate the resumes of individuals at any other company. Online information about both companies and employees allow a recruiter from one firm to search online for candidates from other firms that the recruiter's firm needs. These resumes can then be used to access a company's database and find information about all employees (Cappelli, 2001). So the passive job seeker can be located through information available on the Internet and have offers made to him/her while employed.

Occasionally, companies incur non-value added costs when using online recruiting methods. The costs involved in using a job board as a recruitment tool can be divided into the price charged by the job boards, and the costs incurred by the company in sorting and evaluating the candidates obtained from the use of a job board. To post an ad on a job board such as E-Span, a company can pay anywhere from $150 for a single ad to $4,000 for unlimited advertising in a year. Companies also have the options of creating pages on sites such as Monster.com and Career Mosaic that can cost as much as $10,000 a year (Greengard, 1995). Some boards offer companies the option to pay on a per resume received basis, for a cost of $0.75 a resume. Due to the large volume of resumes that can be received through a job board, a company must contribute additional resources to the sorting and evaluating of the information received (Wyld, 2005). Companies must also spend time competing with other companies when using job boards due to the high availability of access by an individual applicant (Cappelli, 2001). Although the use of job boards can represent additional costs to a company, the availability of qualified applicants and cost savings achieved by online recruiting generally outweigh any costs incurred.

Another potential non-legal problem involved with conducting a job search using electronic resumes submitted through job boards and company websites is the reluctance of job seekers to post their resumes. One quarter of the online job seekers surveyed by Feldman and Klaas (2002) expressed concerns about the security of their personal information on the Internet. To alleviate fears of potential applicants many companies include privacy statements on their websites. Security companies can be used to further reduce fears. One such company is TRUSTe, which provides a "seal" that guarantees the privacy of both personal information and purchases made online (Thomas & Ray, 2000). The presence of such security features can make applicants more likely to submit a resume to an online site.

Relying on a company's website to attract the "passive" job seeker also has its drawbacks. Since this method requires that a potential recruit be on the website, the company's image as well as business will affect the number of visitors (Thomas & Ray, 2000). This can work to the detriment of small companies. A study by Feldman and Klaas (2002) found that the amount of online jobs searches was influenced by the level of an applicants' Internet fluency. Therefore, positions likely to be filled by persons not comfortable or proficient with the Internet will not be filled by relying on the passive job seeker to use the website method.

Legal issues must be addressed when a company is engaged in online hiring. One consideration is disparate impact in the selection process. If a member of a protected category is eliminated from consideration at a greater proportion than a majority category, then disparate impact might be occurring. Although the Internet should, in theory, make job openings available to a greater diversity of candidates, it may in fact discriminate against protected groups (Wyld, 2005).

The ability to access jobs on the Internet may discriminate against protected groups based upon access to the Internet. A study by the National Telecommunications and Information Administration profiled Americans with access to telecommunications equipment. They found that minority households were less likely to have access to computers than white households. Also intercity households were less likely than rural households to have access to computers. Minority households with computers were found less likely to have modems and access to the Internet (Wyld, 2005). Age is also shown to affect Internet access, with individuals over the age of 55 the least likely to use computers (Hogler et al, 1998; Flynn, 2002).

Another legal issue facing online recruiting is the legality of the key-word search. Because the use of a key-word search is a selection tool, it must be valid to be legal. Mohamed, Orife, & Wibowo (2001) proposed three ways that a key-word search could be found to be invalid and therefore illegal. The first would be choosing keywords that may not be related to the job for which the resumes were submitted. A second problem arises with inconsistency of keyword choices. Consistency is necessary to establish validity and, therefore, a company that uses an inconsistent keyword bank may be using a flawed process. The third is related to the resume writing skills of an applicant. In the case where applicants are familiar with prime words used in the keyword search, they would be granted an unfair advantage in the selection process. Flynn (2002) speculated that a keyword search might be used to screen out members of a protected group. Because of the differences in word choices between different races, a company could be using keywords that discriminated against certain groups of people. A lawsuit dealing with such a search against Walt Disney World is wending its way through the courts.

IMPACT ON JOB SEEKERS

Utilizing online recruiting sites affords job seekers with many benefits different than those provided by traditional recruiting methods. Online recruiting is generally speedy, convenient, and inexpensive. It also exposes a vast array of job opportunities to job seekers, and provides them with valuable information about potential employers. Perry (2002) indicates seeking employment opportunities online provide a quick, convenient, reliable, timely, and efficient way to reach employment professionals and organizations without the restriction of geography. Lee, 2005 suggests that submitting electronic applications and resumes allows job seekers to provide professional and personal information almost instantaneously, and they can apply for multiple jobs while submitting their resume only one time.

Online recruiting is convenient for job seekers because access to the Internet is available 24 hours a day and companies will be able to evaluate applicants at anytime. Always having access to testing methods would increase the candidate pool by including people whose schedules would prohibit them from traveling to take a physical test.

Online recruitment is specifically advantageous to the well educated and the computer savvy job seeker. Potential candidates that utilize online recruitment are often highly skilled, and desire demanding and challenging position in their chosen fields. NORAS studies indicate that one in four adults prefer to conduct job searches using web based search engines or other electronic formats. Whether positions are posted on company websites or handled via third party agencies specializing in recruitment, the candidate can locate critical information on the potential employer. A survey by www.wetfeet.com reported that more that 90 percent of job applicants will examine a company's website prior to applying for a position. The job applicants look for hard data including financial status and company reports (McCourt-Mooney, 2000)b, as well as information on the company's culture (Cappelli, 2001). Internet based recruitment can also assist job seekers with resume development tools related to the key areas in specialized fields (Perry, 2002).

Potential problems for job seekers also exist. As more companies and potential employees are using the internet, a number of unanticipated problems occur. Potential employees often experience a lack of information and a slow recruitment process when searching and applying for jobs online. One cause of slow recruitment occurs from the use of data warehouses. Due to the increased amounts of information and resume responses for online postings, many sites have started to use data warehouses to store potential employee information for employer consideration (Maximize Online Recruitment, 2003). In addition, these online employment websites use a 'key word' system that does not always match the best candidates to a particular job (Lin & Stasinskaya, 2002). Also, the speed and efficiency of e-recruiting is frequently lost because hiring managers fail to interview candidates in a timely manner (Make e-Recruiting the Catalyst for HR Systems Integration, 2001). There is also frustration from lack of personal attention and employer feedback (Lin & Stasinskaya, 2002). With this lack of attention and feedback, there are no guarantees for applicants that someone actually looked at their resumes. Thus, many job seekers consider online resume posting inefficient. (Partners, 2000).

Unclear guidelines for follow-up procedures upon posting resumes online are a frequent complaint by online job seekers (Jennings & Hayes, 2000). It is difficult to set the same rules across the entire online career providers as they all try to differentiate themselves by using different navigation, data storage, and retrieval techniques.

Privacy in online hiring is an important issue that is raising concerns among job seekers (Lin & Stasinskaya, 2002). The major concern is lack of assurance that private information on an individual's application form or resume will not be given or sold to a third party or used for other than hiring purposes. Also with the internet being available to essentially everyone, the privacy of applying for a new job is jeopardized by the ability of someone's current employer being able to see their resume online. Employees are concerned about current employers firing them for looking at other jobs, while employers are concerned about employees exposing company secrets (Lin & Stasinskaya, 2002). Concerns also exists about programs that illicitly download resumes for reposting and can also be used for secondary mailings from marketing sources that sponsor the website, without the website being held responsible. Another potential online danger is identity theft. The Identity Theft Resource Center reports that job seekers may be at a slightly higher level of risk for identity theft (Foley, 2005).

A final concern for online job seekers is discrimination. Online job seekers may not have equal or fair opportunities to discover potential job offerings with companies that use online recruiting and hiring methods exclusively to attract potential employees. This is especially true for those in protected groups with limited access to computers and the internet, little knowledge regarding use of computers, lack of knowledge regarding language and writing skills, or unfamiliarity with prime words used in key-word searches. Studies have shown (Flynn, 2002) that younger people are more likely to use online recruitment sites than older people and fewer minorities have access to the internet compared to white people. Feldman & Klaas (2002) found that the amount of online jobs searches was influenced by the level of an applicants' Internet fluency. Also, online job seekers may unknowingly be discriminated against by companies that use filter devices or key-word searches in their recruiting practices (Flynn 2002).

RESEARCH METHODOLOGY, PROCEDURES AND PARTICIPATION

The purpose of this study was to research the perceptions and behaviors of college students towards online recruiting. In order to examine issues related to online hiring and job searching, a survey was conducted with 186 business students to determine who was online and how they were using the Internet in a job search. The survey was adapted from one created by Enhance Media where job boards evaluated the users of their websites. Demographic features included ethnicity, gender, age, and work experience. Comfort level in using online job search technology as well as self-evaluation of computer expertise and use were also captured.

Using a stratified sampling method, several classes were identified and registered students in the course were administered the survey instrument. Participation was voluntary and participants were presumed to possess a working knowledge of online recruiting, specifically in the process and their perceptions of this activity. The survey was conducted anonymously; no personal information was collected that could be used to identify any individual participants.

DEMOGRAPHICS

Participants in the survey are described by the following demographic information. The gender of participants was 50% female and 50% male. The age groups were 66% being 18 to 25 years of age; 23% aged 26 to 35; 7% aged 36 to 45;

4% aged 46 and over. Thirty-nine percent reported being employed full time, 39% part-time, 12% unemployed and looking, while 10% did not work and were not looking. Of the respondents who were employed, 40% had 1 to 5 years experience, 25% had 6-15 years experience, and 8% had 16 or more years. Ethnicity of our sample was 64% white, 15% African American, 1% American Indian, 4% Asian, 12% Latino, and 4% other ethnicities.

RESEARCH FINDINGS AND ANALYSIS

Internet Fluency Level

Respondents provided information about their fluency with the Internet. The results showed that the majority of the participants were aware of the importance of the Internet. Participants self-evaluated their level of technological expertise as follows: beginner (9%); medium (15%); intermediate-computer literate and comfortable with technology (65%); and advanced-technology guru (11.3%).

Experiences and Preferences and Online Recruiting

Two questions were used to measure participants' level of experience and preference toward online recruiting. Preferences for various methods of looking for a job were as follows: 23% reported preferring online job boards; 30% preferred online company websites; 30% preferred newspapers, and 15% preferred using an employment or recruiting agency. Forty-seven percent of the sample did not utilize online sites. However, when asked if they would be willing to post a resume online, the following was reported: 62% yes, anywhere (job board or company website); 18% yes, only on a company website; 8% yes, only on a job board; and 11% would not post their resumes.

Citizens and Non-citizens Usage of the Internet for a Job Search

We looked at differences between citizen and non-citizens in our sample. The chi-square test showed a significant relationship ($p=.01$) between citizenship status and use of the Internet to look for a job. Using a z-test for differences in proportions, we found the following differences between citizens ($n=138$) and non-citizens ($n=48$) in numbers of Internet job searches

Table 1: Internet searches			
	Citizens	Non-citizens	p value
Never used Internet	41.3%	66.7%	p = .002
1 or 2 searches	33.3%	20.8%	p = .10
3 to 5 searches	5.8%	6.3%	p = .91
More than 5 searches	19.6%	6.3%	p = .03

The differences between citizens and non-citizens in obtaining an interview for a job found online, for obtaining a job from an online application, willingness to post a resume online, and for comfort using the Internet were not significant. 96.4% of citizens, compared to 93.4% of non-citizens were comfortable using the Internet. 88.4% of citizens would post their resume online, and 89.6% of non-citizens were willing to post online. We found significant differences in self-reported technological expertise between citizens and non-citizens.

Table 2: Technological Expertise			
	Citizens	Non-citizens	p value
Technophobe	0.0%	2.0%	p = .09
Beginner	5.1%	17.0%	p = .01
Medium	11.6%	23.4%	p = .05
Intermediate	70.3%	48.9%	p = .01
Advanced	13.3%	6.4%	not significant

Of the 186 students surveyed, 95 were actually looking for a job. For the 95 looking, the percentage of non-citizens who never used the Internet was significantly higher than the percentage of citizens who had never used the Internet.

Table 3: Internet Searches			
	Citizens (n=64)	Non-citizens (n=31)	p value
Never used Internet	26.6%	51.6%	p = .02
1 or 2 searches	40.6%	29%	p = .27
3 to 5 searches	10.9%	9.7%	p = .85
More than 5 searches	21.9%	9.7%	p = .15

JOB SEEKERS BY GENDER

Of the 95 job seekers in the survey, 45 were male, and 50 were female. The results according to gender for the number of Internet job searches were:

Table 4: Searches by Gender			
	Men (n=45)	Women (n=50)	p value
Never used Internet	35.6%	34%	p = .87
1 or 2 searches	48.9%	26%	p = .02
3 to 5 searches	6.7%	14%	p = .25
More than 5 searches	8.9%	26%	p = .03

Differences between men and women concerning applying online, obtaining an interview from online application, receiving a job from online, comfortable using the internet, willingness to post resume online, and level of self-evaluated technological expertise were not significant. A significantly higher proportion of men (42% to 24%, p=.06) reported 3 to 5 hours of computer usage while more women (women 18%, men 11%, p=.01) had more than 5 hours of computer usage per day.

JOB SEEKERS GROUPED BY WHITE OR MINORITIES

Of the 95 job seekers, fifty were white, twenty were African American, thirteen were Latino, five were Asian, one was American Indian, and six were classified as other ethnic groups. The 45 nonwhite job seekers were classified as minority, and their responses were compared to those of the white job seekers. There was no significant relationship between ethnicity and number of online job searches. On questions concerning applying for a job online, obtaining an interview from online application, receiving a job from the internet, comfortable using the internet, and willingness to post resume online, there was a significant difference only on applying for a job with 36% white applying and 58% of minority applying, (p = .03). Differences in computer usage between white and minority students were not significant.

For self-evaluated levels of technological expertise, several differences were significant. White students tended to assess their technological expertise at a higher level than the minority students.

Table 6: Technological Expertise by Gender

	White (n=50)	Minority (n-45)	p value
Technophobe	0%	2.2%	p = .29
Beginner	0%	8.9%	p = .03
Medium	10%	28.9%	p = .02
Intermediate	80%	53.3%	p = .01
Advanced	10%	6.7%	p = .50

AGE AND ONLINE JOB ACTIVITY

Of the 95 people who were either actively or passively searching for a job, 67 were between 18 and 25 years old, and 28 were over 25. Job seekers over 25 years old appeared to be more likely to use the Internet to search for a job. The percentages using the Internet to look for a job according to age were:

Table 7: Age and Internet Searches			
	Age 18-25 (n=67)	Over 25 (n=28)	p value
Never used Internet	43.3%	14.3%	p = .007
1 or 2 searches	38%	36.8%	p = .54
3 to 5 searches	9%	14.3%	p = .44
More than 5 searches	9%	39.3%	p = .0004

Differences in age on those answering yes to questions concerning applying for a job on the internet, obtaining an interview from online application, receiving a job from online, and researching a company on the Internet were all significant. In four of these six areas, the older students reported significantly more use of the Internet than those in the 18 to 25 year old group.

For level of technological expertise, the only significant difference was that a larger proportion of the older students considered themselves advanced.

Table 8: Age and Online Recruiting			
	Age 18-25 (n=67)	Over 25 (n=28)	p value
Applied	34.3%	75%	p = .0003
Interview	13.4%	46.4%	p = .0005
Got job	9%	28.6%	p = .014
Researched	52.2%	92.9%	p = .0002
Comfortable	94%	96%	p = .187
Posting online	89.6%	89.3%	p = .97

WORK EXPERIENCE

Of the 95 job seekers, 66 had 5 years or less work experience, and 29 had more than 5 years job experience. The more experienced job seekers appeared to be more likely to use the Internet to search for a job. The percentages using the Internet to look for a job according to work experience were:

Table 9: Work Experience and Searches

Experience	0-5 yrs. (n = 66)	Over 5 yrs.(n = 29)	P value
Never used Internet	39.4%	17.2%	p = .03
1 or 2 searches	36.4%	37.9%	p = .88
3 to 5 searches	9.1%	13.8%	p = .49
More than 5 searches	12.1%	31%	p = .03

Significant differences as far as work experience were not found in applying for a job online, obtaining an interview from online application, receiving a job from online, researching a company online, comfortable using the Internet, and willingness to post a resume online. None of the differences in computer usage between those with more or less work experience were significant.

DISCUSSION AND FUTURE DIRECTIONS

The issue raised in the literature concerning the safety of posting resumes online seems to be of little concern. Of the people surveyed, 89 percent would be willing to post their resumes somewhere online. This suggests that as people are learning more about the Internet and become more comfortable with it (96 percent of respondents) their perception of security issues regarding personal information on the Internet decreases.

When using the Internet to apply for jobs, posting resumes on company sites were more prevalent than on job boards. The company website was also found to be the most popular place to search for a job over newspapers, job boards and recruiting agencies respectively. This, combined with the fact that over half of the respondents researched a company via its website suggests that a company's website is an invaluable tool in the online hiring process. It should also be noted that of the people who visited company's websites, most of them had a more positive image of the company after viewing this site.

Moving hiring processes online allows companies to find and evaluate more candidates at a much lower cost than other processes. When moving online, companies must take care to ensure that their posting and evaluating methods do not have a negative impact on protected groups. The results of the survey suggest that companies who are seeking candidates with at least some college education may not face the dangers associated with discrimination in the areas of online fluency, unfamiliarity with prime words in key word searches, and access to computers. In our study, females were comfortable with and applied online with the same frequency as males. We found that a higher proportion of women than men had conducted more than five Internet job searches and used the computer more than five hours daily. Although whites in our sample assessed their technological expertise at a higher level, a higher proportion of minority respondents had applied for a job online. Participants over the age of 25 had applied for a job online, gotten an interview or a job from an online application, and had researched a company online. However, there were no participants in our sample over 55 years old, and only four in the 46-55 years age group.

If companies choose to participate in online hiring, their most efficient tool will be their own website. A company's website must also be viewed as a tool for promoting the company even if that company does not wish to engage in online hiring. Future research needs to be conducted that would include the groups not represented in the survey findings. Any future research should focus on the use of the Internet as a job search tool and not on household access to computers, as they may return different results.

REFERENCES

Agnvall, Elizabeth (2005). Recruiting by ones and zeros. Retrieved June 3, 2006, from the Society of Human Resource Management Website: http://www.shrm.org/hrtx/library_/nonIC/CMS_014953.asp

Backhaus, K. B. (2004). An exploration of corporate recruitment descriptions on Monster.com. *The Journal of Business Communication*, 41(2), 115-127.

Brice, T.S. & Waung, M. (2002). Web site recruitment characteristics: America's best versus America's biggest. *SAM Advanced Management Journal*, 67(2), 4-8.

Cappelli, P (2001). On-line recruiting. *Harvard Business Review*, 79, 139-147. Retrieved May 12, 2005, from EBSCOhost database.

Co.-specific career web sites: Cost-effective and a good client relationship tool. (2003). *Human Resources Department Management Report.* Retrieved June 3, 2006, from Business Source Premier database, Jacksonville State University Library.

Enhance Media (2003). The National Online Recruitment Audience Survey. Retrieved May 12, 2005, from the World Wide Web: http://www.noras.co.uk/questions/index.html

Feldman, D.C., & Klaas, B.S. (2002). Internet job hunting: A field study of applicant experiences with on-line recruiting. *Human Resource Management*, 41, 175-192. Retrieved May 13, 2005, from EBSCOhost database.

Flynn, G. (2002, April). E-recruiting ushers in legal dangers. *Workforce*, 81, 70-73. Retrieved May 12, 2005, from EBSCOhost database.

Foley, L. (2005). *Fact Sheet 121: Identity theft prevention for job seekers*. Retrieved June 2, 2006, from the Identify Theft Resource Center Website: http://www.idtheftcenter.org/Factsheet121.pdf.

Gale, S.F. (2001, December). Internet recruiting: Better, cheaper, faster. *Workforce,* 80, 74-78. Retrieved May 12, 2005, from EBSCOhost database.

Greengard, S. (1995, July). Catch the wave as HR goes online. *Personnel Journal*, 74, Retrieved May 12, 2005, from EBSCOhost database.

Hadass, Y. S. (2004, February 10). The effect of Internet recruiting on the matching of workers and employers. *Harvard Business Review*, 1-36.

Hamilton, B. (2006) Internet becomes primary hiring source. *Expansion Management*, 21,(3), 4.

Hogler, R.L., Henle, C., & Bemus, C. (1998). Internet recruiting and employment discrimination: A legal perspective. *Human Resource Management Review*, 8, Retrieved May 12, 2005, from EBSCOhost database.

Institute of Management and Administration (2000, March). Online recruiting: What works, what doesn't. *HR Focus*, 77, 5. Retrieved May 12, 2005, from EBSCOhost database.

Jansen, B., Jansen, K., & Spink, A. (2005). Using the web to look for work. *Internet Research*. 15(1), 49-66.

Jennings, A., T., & Hayes, M. (2000). Landing a job in a strange new world. *Journal of Accountancy*, 190(6), 55-61.

Karakanian, M. (2000). Are human resources departments ready for e-hr? *Information Systems Management*, 17, 35-40. Retrieved May 17, 2005, from EBSCOhost database.

Koong, K.S., Liu, L.C., & Williams D.L. (2002). An identification of Internet Job Board attributes. *Human Systems Management*, 21, 129-135. Retrieved May 17, 2005, from INFOTRAC database.

Lee, I. (2005). Evaluation of Fortune 100 companies' career websites. *Human Systems Management*, 24(2), 175-182. Retrieved June 3, 2006, from the Business Source Premier database, Jacksonville State University Library.

Lin, B. & Stasinskaya, V.S. (2002). Data warehousing management issues in online recruiting. *Human Systems Management*, 21, 1-8.

Making e-recruiting the catalyst for hr systems integration. (Sept. 2001). *IOMA's Human Resource Department Management Report*, 1, 10-11).

Maximize online recruitment. (Jan/Feb. 2003). *Strategic HR Review*, 2(2), 5.

McCourt-Mooney, M. (2000a). Internet briefing. *Journal of Managerial Psychology*, 15, Retrieved May 17, 2005, from EBSCOhost database.

McCourt-Mooney, M. (2000b). Internet briefing part II. *Journal of Managerial Psychology*, 15, 512-517. Retrieved May 17, 2005 from EBSCOhost database.

McCourt-Mooney, M. (2000c). Internet briefing part III. *Journal of Managerial Psychology*, 15, 737-740. Retrieved May 18, 2005 from INFOTRAC database.

McCune, J.C. (1998, April). A few good employees. *Management Review*, 87, 38-41. Retrieved May 17, 2004 from EBSCOhost database.

McLean, C. (2006). A foot in the door: IT. *Certification Magazine*. 8(4), 38-40.

Menagh, M. (1999, January). IT cost per hire: finding net (and other) savings. *Computerword*, 50. Retrieved May 30, 2004, from INFOTRAC database.

Minton-Eversole, T. (2006). Best Online Source for Hiring Can Be in Company's Own Virtual Backyard. Retrieved June 3, 2006, from the Society of Human Resource Management Website: http://www.shrm.org/ema/news_published/CMS_015835.asp

Mohamed, A.A., Orife, J.N., & Wibowo, K. (2001). The legality of key word search as a personnel selection tool. *Employee Relations*, 24, 516-522. Retrieved May 18, 2005, from Emerald database.

Monster Advantage-May 2006. (2006). Monster.com. Retrieved June 3, 2006, from http://media.monster.com/a/i/infomons/pdf/mayadvantage.pdf

Monster Fact Sheet. (2006). Monster.com. Retrieved June 3, 2006, from http://library.corporate-ir.net/library/13/131/131001/items/171946/factsheet1005.pdf

Mooney, J. (2002). Pre-employment testing on the Internet: Put candidates a click away and hire at modem speed. *Public Personnel Management*, 31, 41-53. Retrieved May 12, 2005, from EBSCOhost database.

Partners, Y. (2000). Personal touching is 'missing link' in online recruiting. *The CPA Journal*. 70(2), 8.

Perry, Phillip., (2002) Battle for the best: What works today in recruiting top talent. *Research-Technology Management*. 4(2), 1-8.

Quible, Z.K. (1998). The electronic resume: an important now job-search tool. *Journal of Education for Business*, 74, 79-83. Retrieved May 17, 2005, from EBSCOhost database.

Robb, D. (2004, April). Career portals boost online recruiting. *HR Magazine*, 49, 111-115. Retrieved May 12, 2005, from EBSCOhost database.

Schweyer, Allan. (2003). *Talent management systems*, New York: Wiley, 2003.

Thomas, S.L., & Ray, K. (2000). Recruiting and the web: high-tech hiring. *Business Horizons*, 43, 43-53. Retrieved May 17, 2005, from EBSCOhost database.

Volpe, L., & Tucker, J. (2004). Third-party recruiting. *Employment Relations Today,* 31, Retrieved May 18, 2005, from Wiley InterScience database.

Wyld, D.C. (2005). Bits and paper: The emerging employment market in cyberspace. *American Business Review*, 16, 64-71. Retrieved May 17, 2005, from EBSCOhost database.

Zall, Milton. (2000). Internet recruiting. *Strategic Finance*. 81(12), 1-6.

IMAGE COMPRESSION AND FEATURE EXTRACTION USING KOHONEN'S SELF- ORGANIZING MAP NEURAL NETWORK

Dinesh K. Sharma, University of Maryland Eastern Shore
Loveleen Gaur, BLS Institute of Management
Daniel Okunbor, Fayetteville State University

ABSTRACT

Data transmission over the Internet is prevalent and the development of efficient algorithms for compressing such data in order to achieve reduced bandwidth has been an active research. With increased demand for exchanges of audio, video and other image data over the Internet, research for data compression is more intense than ever before. One of the most popular algorithms for image compression and feature extraction is the use of Kohonen's Self- Organizing Map (SOM), a neural network scheme, designed originally for pattern recognition through association rules. In the past decade and half, the SOM has revolutionized the field of data compression and feature extraction because of its ability to reduce higher dimensional data into lower dimensional data. In this paper, a global processing technique is proposed for training the Kohonen's network and this new technique was tested using JPEG images and the compressed images were considerably reduced in size.

Keywords: Kohonen's Self-Organizing Map, Feature extraction, Image compression, Global processing, Neural Network.

INTRODUCTION

The rapid development of information and communication technologies is enabling large amount of information to be processed, stored, and transmitted over high speed networks. The need for data compression and transmission is increasingly becoming a significant topic in all areas of computing and communications. Computing techniques that would considerably reduce the image size that occupies less space and bandwidth for transmission over networks form an active research. Image compression deals with reducing the amount of data required to represent a digital image (Haykin, 2003).

Digital image processing: (1) improves the information content of the picture for better understanding of what the picture is made up of so that an image appears to be better as its contrast is increased; (2) provides data compression for efficient storage and transmission. Although, storage technology has improved significantly over the past decade, the same cannot be said for transmission capacity. Recently, the need for the transmission of the visual data has increased, particularly over the Internet. Image compression is a commonly used practice in Joint Photographic Experts Group (JPEG) image compression standard. The JPEG image files generally have the jpeg in their name extensions.

Several compression techniques have been developed, such as Differential Pulse Code Modulation, Discrete Cosine Transform (Gersho and Gray, 1992), Discrete Fourier Transform, and numerous Vector Quantization (VQ) methods (Amerijckx et. al., 1998). VQ technique has been used widely (Gray, 1984; Gersho and Gray, 1992). VQ has fairly good performance in both compression ratio and extracted image quality. The principle of the VQ technique is simple. At first, the image is split into square blocks of n´n pixels, for example 4´4, 6´6 or 8´8; each block is a vector. After dividing the original image into blocks, the VQ encoder is used to search each block throughout the codebook for the codeword that is similar to the image block. The index values of the code words closest to the blocks are recoded as the compressed image. When decompressing the image, the VQ decoder uses these index values to recover the corresponding blocks to reconstruct the image. Although, VQ is a powerful technique, it however suffers from high computational complexity.

SOM is another image compression technique that achieves the same efficiency as that of the VQ scheme (Amerijckx et. al., 1998). SOM has many applications, including but not limited to, classification and exploration of collections of text documents, image compression and pattern recognition (Chen et al., 1994; Pei and Lo, 1988). Kurnaz et al. (2001) presented an incremental SOM for the segmentation of ultrasound images. Elements of the feature vectors were formed by the Fast Fourier Transform of image intensities in square blocks. Zheng (1994) described the concept of groupings based on certain rules such as proximity and similarity of image segmentation based on SOM. Katoh et al. (1998) used SOM for recognizing human emotions through facial expressions. By inputting various images of facial expressions into SOM and changing the interconnection weights, they were able to classify image features.

In this paper, we present an image compression technique using SOM employing global processing. Global processing technique processes every pixel of an image without utilizing blocks as it is generally implemented in the conventional SOM. The proposed technique places emphasis on the fourth property of the feature map, e.g., feature extraction. In this respect, the data from an input space obtained from nonlinear distribution, is processed using the SOM to select the set of best approximating features.

The remainder of the paper is organized as follows. In Section 2, a brief review of relevant literature is presented. In Section 3, the methodology used in the paper is described. In Section 4, an application of the methodology is offered. In Section 5, the result of the methodology is presented. In Section 6, conclusions are given.

BACKGROUND

Neural Networks

Neural networks have changed the way we solve "real-world" problems in science, engineering and economics (Mennon, 1996). Neural networks are an effective tool for pattern classification and clustering (Haykin, 2003). There are broadly two paradigms of neural learning algorithms namely, supervised and unsupervised. In the supervised learning paradigm, the networks are generally universal approximations of continuous/discontinuous functions, with prior knowledge of the information about the input-output map. Such known information is used to train the network, combined with backward/feed forward propagation to obtain the optimal interconnection weight matrix. The trained network is utilized for subsequent simulations. The quality of output from the supervised network is determined by its closeness to the desired actual output. The unsupervised neural network is the opposite of supervised network, in this case, desired actual output is unknown. Unsupervised neural algorithms are best suited for clustering patterns on the basis of their inherent characteristics (Kohonen, 2001; Haykin, 2003). The major approaches for unsupervised learning: a) Competitive Learning, b) Self Organizing Feature Maps.

Feature Extraction

Feature extraction is the process of mapping the original features (measurements) into fewer features which include the main information of the structure of the data. Feature extraction methods can be grouped into four categories based on *a priori* knowledge: supervised *versus* unsupervised; and by the functional form: linear *versus* nonlinear. In cases where the target class of the patterns is unknown, unsupervised methods are the only way to perform feature extraction. In other cases, supervised paradigms are preferable. Linear methods are simpler and are often based on an analytical solution but they are inferior to nonlinear methods when the classification task requires complex hyper surfaces. Widespread unsupervised methods for feature extraction are Principal Component Analysis (PCA), a linear mapping; and Sammon's nonlinear mapping. The PCA attempts to preserve the variance of the projected data, whereas Sammon's mapping tries to preserve the interpattern distances.

Feature extraction can be seen as a special kind of data reduction in which the goal is to find a subset of informative variables based on image data. Since image data are by nature high dimensional, feature extraction is often a necessary step for segmentation or object recognition to be successful. Feature Extraction is the first and most important step which is ordinarily performed in an unsupervised manner for pattern classification. The objective of this step is to select small sets of features in which the essential information content of the input data is concentrated. The property "feature selection" of SOM is well suited for the task of feature extraction particularly if input data is generated by a non-linear process (Haykin, 2003).

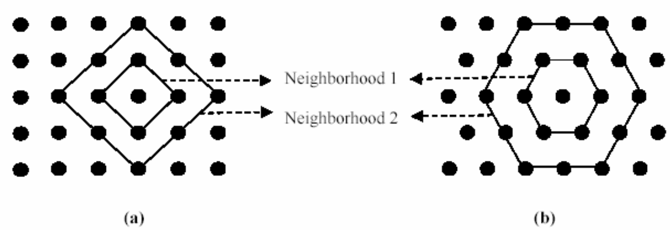
Kohonen's Self-Organizing Maps

Kohonen invented the Self-Organizing Map (SOM) in the early 1980s. Kohonen's SOM is a widely-used artificial neural network (ANN) model based on the idea of self-organized or unsupervised learning (Kohonen, 2001). The SOM network is a data visualization technique, which reduces the dimensions of data through a variation of neural computing networks. It is a nonparametric approach that makes no assumptions about the underlying population distribution and is independent of prior information (Kohonen, 2001). The problem that data visualization attempts to solve is that humans simply cannot visualize high dimensional data so techniques must be created to help us understand high dimensional data.

In comparison, the back propagation algorithm requires the examples to consist of input-output pairs. The network architecture of the SOM consists of a set of laterally interacting adaptive processing elements, nodes, usually arranged in a two-dimensional grid called a map. All the map nodes are connected to a common set of inputs. Any activity pattern on the input gives rise to excitation of some local group of map nodes. After learning, the spatial positions of the excited groups specify a mapping of the input onto the map. The learning process is based on similarity comparisons in a continuous space. The result is a system that maps similar inputs close to each other in the resulting map. The input may be highly complex multidimensional data like in real-life speech recognition, image analysis, and process monitoring.

SOM goes about reducing dimensions by producing maps of usually one or two dimensions which plot the similarities of the data by grouping similar data items together. So SOM's accomplish two things, they reduce dimensions, and display similarities. SOM is a type of unsupervised learning where the goal is to discover some underlying structure of the data. SOM is called a topology preserving map because there is topological structure imposed on the nodes in the network. A topological map is simply a mapping that preserves neighborhood relations. The neighborhood can be rectangular or hexagonal (Figure 1).

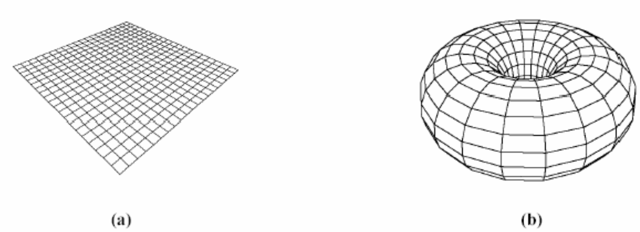
Figure 1: The SOM grid structure: (a) rectangular (b) hexagonal



In SOM, the neurons are placed at the nodes of a lattice, e.g., usually one or two dimensional. The neurons become selectively tuned to various input patterns (Stimuli) or classes of update patterns in the course of a competitive learning process. The location of the neurons so tuned (i.e. winning neurons) become ordered with respect to each other in such a way that a meaningful coordinate system for different input features is created over the lattice. A SOM is therefore characterized by formation of a topographic map of the input pattern in which the spatial location of the neurons in the lattice is indicative of intrinsic statistical features contained in the input pattern (Haykin, 2003; Zheng, 1994).

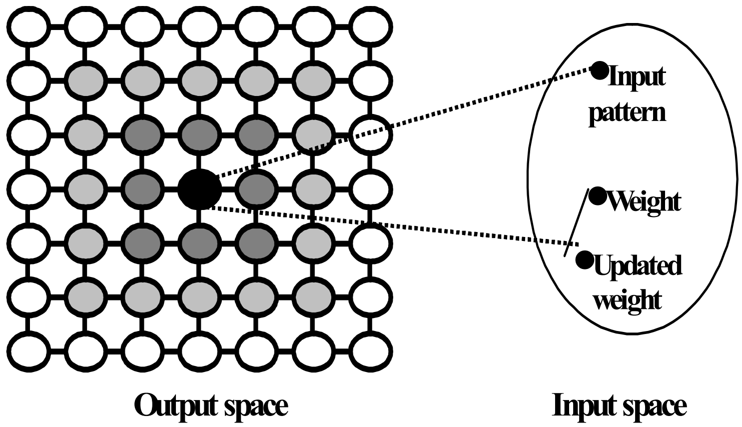
SOMs are based on competitive learning in which the output neurons of the network compete among themselves to be activated or fired, with the result that only one output neuron, or one neuron per group, is on at any one time. An output neuron that wins the competition is called a winning neuron (Haykin, 2003). One way of inducing a winning neuron is to use lateral inhibitory connections (i.e., negative feedback paths) between them. SOM provides a way of representing multidimensional data in much lower dimensional spaces - usually one or two dimensions. The brain is organized in such a way that topologically ordered computational maps (defined by an array of neurons representing slightly differently tuned processors or filters) represent different sensory inputs. Consequently, the neurons transform input signals into a place-coded probability distribution that represents the computed values of parameters by sites of maximum relative activity within the map (Knudsen, 1987).

Figure 2: Two types of SOM topologies (a) planar (b) toroidal



From the above Figure 2, it can be seen that neurons can be placed in different topologies. Besides the topologies shown above there is a random topology, in which the neurons are placed randomly.

Figure 3: The basic architecture of the SOM



Kohonen's SOM Algorithm

A SOM does not need a target output to be specified unlike many other types of networks. Further, when the node weights match the input vector, that area of the lattice is selectively optimized to more closely resemble the data for the classification of the input vector. Kohonen's network follows the "Winner takes all" policy. The network cluster unit, whose weight vector matches more closely with the input pattern, is considered the winning neuron. The weights of the winning unit and its neighboring units are updated (Figure 3). The winning unit is decided based on the Euclidean distance (D) is calculated as:

$$D = \sqrt{\sum} (x_i - w_{ij})^2$$

During the training process, input data is fed to the network through the processing nodes in the input layer. Training occurs in several steps and over several iterations. The following training procedure summarizes the Kohonen learning algorithm ((Haykin, 2003; Kohonen, 2001).

Step 1. Initialization : Assign randomly small values to the initial weight vectors $W_j(0)$ of output neuron j, Where $j = 1, 2. . . N$.

Step 2. Sampling: Draw an input vector X randomly from the input data, and feed it into the Network.

Step 3. Similarity Matching: Find distance between input vector X and each output neuron's weight $W_j(n)$ at time n.

$Dj = ||X - Wj||; j = 1, 2. . . N,$ where $||.||$ is Euclidean distance.

Step 4. Selecting: Select the winning neuron C which has minimum of Dj.

$C = \arg j \min (Dj); j = 1, 2, . . ., N.$

Step 5. Updating: Adjust the weight vectors of all neurons through

$wj (n + 1) = wj (n) + (n)[X(n) - Wj (n)]$ if $\in Ac$
$wj (n + 1) = wn(n)$ otherwise
$C = \arg j \min (Dj); j = 1, 2. . . N,$

Where, (n) is the learning rate parameter, and AC is the neighborhood function centered on the winning neuron C.

Step 6. Continuation: Repeat steps (2)–(5) until no noticeable changes in the feature map are observed or when specified number of epochs is achieved (Haykin, 2003).

METHODOLOGY

As mentioned earlier, the Kohonen's SOM is based on an unsupervised learning that only requires the input data with the main objective of reducing high dimensional input space to lower dimensional output. In addition to these important characteristics, the algorithm is simple and easy to understand. In this paper, we use SOM for segmenting images. Image processing using segmentation is treated as a classification problem, in which segmentation is achieved by pixel classification using SOM (Kong and Guan, 1994). Kong et al. (2002) used SOM for performing image segmentation in two steps, coarse segmentation to obtain the global clustering information of the image followed by pixel based classification scheme that utilizes the local features to refine segmentation. Iivarinen et al. (1996) used SOM to estimate the distribution of features extracted from faulty-free samples. Visa (1992) implemented image segmentation based on SOM and texture measures. The proposed SOM algorithm for image compression using SOM employing global processing is divided into six steps as depicted in the Figure 4.

Figure 4: Flowchart of the methodology

Step 1- Fixing the objectives, namely creation of networks (various dimensions), 3-dimensional matrix formation for the input

↓

Step 2- Reducing the jpeg digital image into 100X100 and creation of 3D matrix form which will be used as Input

↓

Step 3- Different dimensional Networks are created using Network/Data Manager tool of MATLAB like 20X20, 25X25, 35X35,…, and so on

↓

Step 4- In this phase 3D matrix was given as Input to the Networks created in Phase2 where node weights match the input vector which is termed Training of Networks

↓

Step 5 –After much iteration SOM makes stable zones which are graphical outputs

↓

Step 6–After the completion of Training weights are extracted and images are reconstructed based on these weights.

Networks were created to compress the image while retaining its key features. Experiments were done with the help of Network/Data manager tool of the MATLAB software. Every network created was uniquely identified by its name. The networks created were Self-Organizing Maps having dimensions as 15x15, 20x20, 25x25, 35x35 and finally 45x45 dimensional map varied from 225 neurons to 2025 neurons. The topology adopted was a grid topology. The distance function selected was 'LINKDIST'. The ordering phase learning rate was fixed at 0.9 and the steps for this phase were kept at 2000. Similarly tuning phase learning rate was kept at 0.02. The neighborhood distance was fixed to unity.

After the creation of the network, a three-dimensional matrix (as described above) was given input into the network. Where the node weights match the input vector, that area of the lattice was selectively optimized to more closely resemble the data for the classification of the input vector. From an initial distribution of random weights, and over much iteration, the SOM eventually settle into a map of stable zones, where each zone was effectively a feature classifier, so you can think of the graphical output as a type of feature map of the input space. Self-Organizing Feature Maps (SOFM) learn to classify input vectors according to how they are grouped in the input space. They differ from competitive layers in that neighboring neurons in the self-organizing map learn to recognize neighboring sections of the input space. As a result, self-organizing maps learn both the distribution (as do competitive layers) and topology of the input vectors they are trained. After the completion of training, weights were extracted from the network object, obtained as a result of training. The input/output algorithms using MATLAB coding are given in Figures 5 & 6.

Figure 5: INPUT Algorithm using MATLAB coding

```
Step 1: Input jpeg image
            a1= imread ('image',' jpeg');
            a1=double (a1);

Step 2: Reduce the jpeg image into 100X100 pixels
            a1=a1 (1:100, 1:100);

Step 3: Initialize variables l=100, m=100,a=[0 0 0], c=[0 0 0] c=[c` ,a`]
            For i =1: l
              For j=1: m
                  a= [i, j, a1 (i,j)];
                   a=[c a'];
                   c=a;
                   end
              end
            c=c(:,3:10002):
Step 4: c is the Output which is a matrix form and work as input for networks
```

Figure 6: OUTPUT Algorithm Using MATLAB Coding

```
Step 1: load filename;        (Input is Network of various dimensions like 15X15, 20X20,..,
                               and, so on.)

Step 2:   w=network9.iw;
           a=0;
          ww = cell2mat(w);
            a=uint8(a);
          ww=uint8(ww);
          For n=1:1225
             i=ww(n,1);
             j=ww(n,2);
             k=ww(n,3);
             a(i,j)=k;
          end
```

(Here the node weights match the input vector, that area of the lattice is selectively optimized to more closely resemble the data for the class the input vector is a member of. From an initial distribution of random weights, and over much iteration, the SOM eventually settles into a map of stable zones. Each zone is effectively a feature classifier, so you can think of the graphical output as a type of feature map of the input space)

```
Step 3: imwrite (a, 'test11', 'jpeg'); (Converting these output files into jpeg)
```

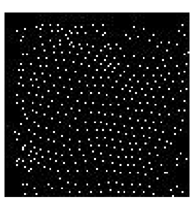
RESULTS

Based on node weights taken from the input algorithm, reconstruction of the image was performed using an output algorithm and the results were found to be satisfactory. We noticed that as we increased the dimensions of the SOM, the images began to resemble more closely the original image. Also, more feature extraction took place as we increased the dimensions of the map. The more closely resembled reconstructed image i.e. 45X45 is reduced to 5409 bytes. Since SOM is computationally intensive, a substantial amount of time is expended in the mere training of any SOM network.

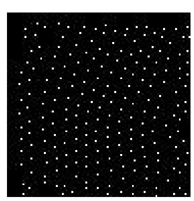
CONCLUSION

In this study, we used SOM for segmenting images. The unsupervised learning paradigm implemented using SOM has been one of the major methods for image segmentation research. Image compression addresses the problem of reducing the amount of data required to represent a digital image. Digital Image Processing encompasses processes whose inputs and outputs are images and encompasses processes that extract attributes from images, up to, and including the recognition of individual objects. A global processing technique was used for the image compression and the image taken was in jpeg format. As we increased the dimensions, the picture was reduced by the number of bytes and started to closely resemble the actual picture through the feature extraction property of SOM thereby making the images very convenient for storage and transmission.

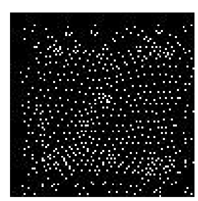
Figure 7: Reconstruction of Images using Output algorithm



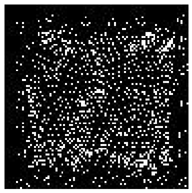
(a) 15X15, 4093 bytes



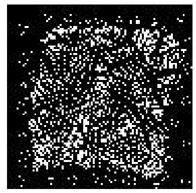
(b) 20X20, 4488 bytes



(c) 25X25, 4692 bytes



(d) 35X35, 5162 bytes



(e) 45X45, 5409 bytes

(f) Original Image = 2835X2050, 5811750 bytes

REFERENCES

Ahmed, N., T. Nataranjan & K.R. Rao (1974). Discrete Cosine Transform. *IEEE Trans. Com.*, C-23, 90-93.

Amerijckx, C., M. Velerysen, P. Thissen & J.D. Legat (1998). Image compression by self-organized Kohonen map. *IEEE Transactions on Neural Networks,* 9 (3), 503–507.

Chen, O.T.-C., B.J. Sheu & W.-C. Fang (1994). Image compression using self organization networks. *IEEE Trans. Circuits Syst. Video Tech*., 4 (5), 480–489.

Delogne, P. & B. Macq (1991). Universal Variable Length coder for an integrated approach to image coding *Ann Telecomm.,* 46(7-8), 452-459.

Gersho, A. & R.M. Gray (1992). *Vector Quantization and Signal Compression*. London: Kluwer Academic Publishers.

Gray, R.M. (1984). Vector qunatization. *IEEE ASSP Mag.*, 9–31.

Haykin, S. (2003). *Neural Networks: A Comprehensive Foundation*. New Delhi: Prentice Hall India.

Iivarinen, J., J. Rauhamaa & A. Visa (1996). Unsupervised segmentation of surface defects. *Proceedings of ICPR*, 356-360.

Katoh, A. & Y. Fukui (1998). Classification of facial expressions using self organizing maps. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in medicine and Biology Society*, 20, 986- 989.

Kohonen, T. (2001). *Self Organizing Maps*. Berlin: Springer-Verlag.

Kong, H. & L. Guan (1994). A self-organizing neural network for image segmentation. *Proceedings of the 1994 Second Australian and New Zealand Conference*, 27-31, 29 Nov.-2 Dec., 1994.

Kong, H. S., L. Guan & S. Y. Kung (2002). A self organizing tree map approach for image segmentation. *Proceedings of ICSP*, 588-591, 2002.

Knudsen, E. I., S. D. Lac & S. D. Esterly (1987). Computational Maps in the Brain. *Annual Review of Neuroscience*, 10, 41-65.

Kurnaz, M. N., Z. Dokur & T. Olmez (2001). Segmentation of ultrasound images by using an incremental self-organized map. *Proceedings of the 23rd annual EMBS International Conference, Istanbul, Turkey*, 2638-2640, October, 25-28.

Macq, B. (1990). A Universal entropy coder for transform on hybrid coding. Picture Coding Symposium, 12.1.1-12.12.2, Boston.

Mennon, A., Mehrotra, C. Mohan & S. Ranka (1996). Characterization of a Class of Sigmoid Functions with Applications to Neural Networks. *Neural Networks*, 9, 819-835.

Pei, S.-C. & Y.-S. Lo (1998). Color image compression and limited display using self-organization Kohonen map. *IEEE Trans. Circuits Syst*. 8 (2), 191–205.

Stallman, Richard (2001). The GNU Project (www.stallman.org/rms-bw.jpeg).

Visa, A. (1992). Unsupervised image segmentation based on a self-organizing feature map and a texture measure. *Proceedings 11th IAPR International Conference on Image, Speech and Signal Analysis,* Vol. III , 101-104, 30 Aug-3 Sept, 1992.

Zheng, Y. J. (1994). Self organizing grouping for feature extraction and image segmentation. International symposium on speech. Image Processing and Neural Networks, 13-16, April, 1994.

Zirilli, J. (1997). *Financial Prediction using Neural Networks*. Boston: International Thompson Computer Press.

E-COMMERCE DISRUPTIVE INNOVATIONS IN CHARITY AND NON-PROFIT FUND RAISING

Chung-Shing Lee, Pacific Lutheran University
Eli Berniker, Pacific Lutheran University
Glenn Van Wyhe, Pacific Lutheran University
Kenneth J. Johnson, Pacific Lutheran University

ABSTRACT

Philanthropic fund raising has evolved into a major business activity. Traditional models utilize one-on-one contact, direct mail, and telephone bank fund raising. Most e-giving and e-philanthropy companies' business models are simply the extension of the traditional brick-and-mortar paper-oriented processes. The transaction costs of these models are extremely high, often representing more than 50% of contributed funds. Neither are these methods very effective. The high transaction costs and limited effectiveness offer an opportune cost umbrella for e-philanthropy, web-based fund-raising systems. Many business models have evolved into a set of opportunities that promise to disrupt traditional fund raising models. This paper examines the question of how online e-giving companies can capitalize on the disruptive attributes of the Internet and e-commerce to redesign their business and revenue models to achieve high performance. Our purpose in this paper is to examine the disruptive potential of e-philanthropy. We will propose value propositions and architecture of processes appropriate for web based non-profit fund raising. We will classify, compare, and evaluate various business models that promise to change the nature of philanthropic fund raising. Overall, the gains for the online company lie within the design and implementation of the virtual value chain and what the firm can do with the information it collects.

Keywords: philanthropy, e-commerce, innovation, charity, fund raising, and business model

INTRODUCTION

There were about 1.4 million charities, social welfare organizations, and religious congregations in the United States in 2004. Total giving to those charitable organizations rose to $248.5 billion USD in 2004, an increase of 2.3 percent from 2003 (when adjusted for inflation). The majority of that giving ($187.9 billion USD) came from individuals.[1] Before September 11, 2001, no single online group raised more the $3 million dollars (Blau, 2001a). A 1999 survey performed in the U.S. reported that 1.2% of total contributions came in over the Internet.[2] Giving steadily increased during the majority of the 1990s, slipping only in 2001. A survey in 2001 found that only 1% of those surveyed donated using the Internet.[3] Since September 11, 2001, relief charities took in online more than $215 million USD of the over $2 billion collected, or 10.8% (Wallace, 2002). One third of what the September 11th fund-raising received came in through the Internet; with an average size of $10 USD more than traditional methods (Wallace, 2002). In 2005, 63% of households in the United States had Internet access. Obviously, the potential for growth in online charity and fundraising is very significant.

In this paper we will use the term "e-philanthropy" to include e-giving, online charity or gift-giving, intranet workplace giving, or online donations; terms that are all more or less synonymous with each other. Austin (2001) describes e-philanthropy as the use of the Internet to raise money and recruit volunteers. The concept allows individuals the ability to set up donation pledges and facilitates the electronic transfer of funds to the charity or organization of one's choosing. Implementing this type of process, or service, could potentially replace time- and resource-consuming processes with an outsourced provider of gift-giving options who can expedite the movement of a donation through electronic means.

This paper first looks at the characteristics and major concerns of the e-philanthropy industry in the United States. Next, the paper argues that the e-philanthropy model is a disruptive innovation that fundamentally changes the traditional approach of gift-giving and charity donation. E-philanthropy business models are then identified and defined. In addition, a new approach to e-philanthropy is introduced in section 5. In closing, some specifics that could add value to this process and service will be identified. Moreover this paper will examine what those in the industry might profitably contemplate when searching for ideas to broaden their market without the infusion of substantial resources. These ideas must take into account what individuals considering this type of service may require and the circumstances surrounding the need before additional service options are produced.

THE E-PHILANTHROPY INDUSTRY

his section discusses issues of concern to the user/donor, and service-providers in the e-philanthropy industry. We will additionally address major characteristics of the industry and how they will affect the success or failure of those involved.

Donor's Concerns

Donors utilizing e-philanthropy want to know their funds are being transferred directly to the charity or non-profit organization (NPO) they choose, and that the information they pass on is secure and private; that they can trust the service or organization and the electronic funds transfer (EFT) process. The Association of Fundraising Professionals has announced the development of principles for the e-donor Bill of Rights.[4] This would give companies and NPOs guidelines that could help alleviate concerns such as transaction security which includes trust and legitimacy issues.

Marketing and Non-Profit Concerns

A new way of thinking is needed in providing an e-philanthropy service. Organizations (both for-profit and non-profit) starting online-giving programs will be far more successful if they offer well-designed websites and innovative marketing (Hruby, Blum, and Voelz, 2001). Another issue is that NPOs have been slow to accept electronic programs, in part because many online-giving websites were founded by persons with no NPO experience (Fix, 2001). Additionally, NPOs have been reluctant to get involved with for-profit companies.[5] The main here issue is: Can NPOs learn from and capitalize on the success of many e-commerce business models and practices, such as providing customized services, to get their "customers" to embrace this new way of giving?

INDUSTRY CHARACTERISTICS

The e-philanthropy industry is influenced by several economic factors that commonly exist in other high-tech or knowledge-intensive industries.

Network Effects

Whether the network effect is at work in the e-philanthropy industry is an empirical question. Nevertheless, it can be understood that people give to what they believe to be successful programs (Blau, 2001a). The more users of an online-giving website, more perceived value from the "customers." Utilizing a service for accepting donations electronically is not a high cost endeavor, nor is updating online information that could reassure potential donors that their money will be properly utilized towards the cause they seek to support. Porter and Karmer (2002) point out that individuals rarely have the time or expertise to undertake such serious due diligence when it comes to uncovering who and what benefits from their donation. Utilizing news and media outlets of all types, the Internet, and the donors network effect, individuals can be quite aware of where their money goes. Charities and NPOs have the ability to educate potential donors on what they do and how they do it. Constantly updated facts and figures, Internet seals of approval (i.e. the trust factor), and customer evaluation dialog and feedback within user communities can all be utilized to help put a cautious donor at ease, fostering strong network effects.

Increasing Returns and Positive Feedbacks

"Increasing returns" occur when an organization which is ahead gets farther ahead, or when it loses advantage and then loses further advantage (Arthur, 1996). The possibility of one or a few participants in the e-giving industry growing rapidly by attracting new supporters or donors who ultimately choose the dominant player(s) in the industry is very likely. If the winner-takes-all scenario plays out in this industry, the Internet could become the beginning of the end for various charities and NPOs. In fact, scale matters in the e-philanthropy industry due to thin margins that make volume crucial for profitability (Austin, 2001). Below-cost pricing is a common strategy to build a critical mass of installed base of users among for-profit e-philanthropy companies. In addition, Austin (2001) points out that smaller players will not necessarily be able to mobilize the capital necessary to cover the marketing costs to drive traffic or the technology investments needed to create innovative online services6[16].

However, the possibility of a winner-take-all scenario coming into play could be greatly diminished by the amount of information an organization makes accessible to those interested. If individuals are able to see how their donations are spent and charities are able to promote how they spend money, a much more level playing field could emerge. Differentiation would come down to what service supports the recipients of the charity in question more efficiently, not necessarily who has the most popular brand name.

Digital Divide

Many NPOs still rely on traditional “brick-and-mortar” styles of collecting and distributing donated funds. The digital divide is present for two reasons: lack of affordability and fearfulness of the technology. The online-giving revolution is rooted in underlying behavioral changes (Austin, 2001), in addition to technological challenges. It is essential to develop technology awareness and Internet readiness for NPOs. NPO managers need to understand the business strategies of for-profit e-philanthropy companies and to assess the opportunity of establishing partnerships with them. The digital divide may grow smaller in time as costs go down and pressure to “get on the Web” overcomes stubbornness and fear.

Competition

Most of the e-philanthropy donation distribution services industry is made up of for-profit companies. However, donation services websites may be discovering that nonprofit organizations have better brand names than they do (Fix, 2001). In addition, they may be finding the initial success of one service provider, or greater visibility of a specific NPO, can generate public perceptions that the provider is the charity of choice (Blau, 2001a). Although network effects and increasing returns could set some organizations apart, the options available to organizations that do not want to be left behind are numerous and the costs are minimal. Utilizing the Internet will allow even the small, local and specialized charities to state their case, inform the public and lay out how their program(s) work. Online giving creates conditions in which nonprofits come into more direct competition for donation dollars (Blau, 2001a). An online provider of the donation distribution processes can put available charities side-by-side; giving any of them a chance to make their case, helping to narrow the divide. This type of contrast and comparison allows those with good and ethical programs to thrive and grow. The profound structural changes made possible by technologies in the nonprofit sector will remake what it means to be an effective nonprofit organization (Blau, 2001b).

E-PHILANTHROPY IS A DISRUPTIVE STRATEGIC INNOVATION

In this section we discuss why the e-philanthropy business model is a disruptive strategic innovation that has fundamentally changed the way of competing in the traditional philanthropy industry.

Disruptive Strategic Innovation

e-philanthropy is an innovation in the Digital Economy, i.e., the use of new knowledge (both technological and market) to offer products or services that customers want. e-philanthropy is not a radical innovation but it is disruptive in that the traditional organizations (e.g., NPOs) lack the necessary models of competitive architecture and organizational capabilities, and are therefore unable in critical ways to do what must be done (Miller and Morris, 1999). Typically, a disruptive innovation presents a different package of performance attributes – attributes that, at least at the outset, are not valued by existing customers (Bower and Christensen, 1995). The performance attributes that existing customers do value improve at such a rapid rate that the new innovation can later invade established markets. In general, disruptive innovations create an entirely new market through the introduction of a new kind of product or service (Christensen and Overdorf, 2000). Charitou and Markides (2003) extend the concepts of disruptive innovation (Bower and Christensen, 1995) and strategic innovation (Markides, 1997) to introduce the concept of disruptive strategic innovation. According to Charitou and Markides (2003), disruptive strategic innovation is a specific type of strategic innovation. It is a new way of competing in the industry that is both different from and in conflict with the traditional way. Online brokerage trading and Internet banking provide two good examples of disruptive strategic innovation.

e-philanthropy provides a new architecture to an existing service. It emphasizes different product or service attributes. With the increase in options for both organizations and users, along with the flexibility, speed, and fee alternatives, it also brings to market a very different value proposition than had been available (Christensen, 1997). In addition, e-philanthropy started out as small and low-margin "businesses." Generally, disruptive strategic innovations underperform established mainstream products or services. As mentioned in the introduction of this paper, only one percent of charitable giving was done through the Internet in 2001. This is true even though the new service model is just as secure as the traditional model and has additional value propositions. Over time, e-philanthropy business models and technology will improve to the extent that they are able to deliver performance considered good enough in the old attributes, that established competitors require, while offering new attributes that produce superior potential, attracting a large share of the established market.

DISRUPTIVE ATTRIBUTES OF E-PHILANTHROPY

Lee (2001) identifies and discusses several disruptive attributes of the Internet and e-commerce, such as open platform, prosumption[7] (Tapscott, 1996), digital assets[8] (Rayport and Sviokla, 1995), virtual capacity[9] (Afuah and Tucci, 2003), information sharing and exchange[10] (Evans and Wurster, 1997), cost transparency [11] (Sinha, 2000), industry scope[12], network connectivity and real-time interactivity, and the speed and frequency of technological and organizational changes. Lee (2001) argues that business executives must be able to capitalize on the key performance attributes of e-commerce innovation to improve overall business performance. He defines business model innovation in e-commerce as the use of new knowledge (both technological and market) that capitalizes on the disruptive attributes of the Internet to design and implement an innovative way of offering products or services that customers want. To understand the disruptive nature of an innovation or technology, one must first identify the unique attributes that make it disruptive to the traditional organizations or established markets. The following two functions can be applied to assist business executives or entrepreneurs in identifying and capturing the benefits of a disruptive innovation.

$$T_S\,(X_1,\ldots,X_n;\,Y_1,\ldots,Y_m)$$
$$T_D\,(X_1,\,\ldots,X_n;\,Z_1,\ldots,\,Z_l)$$

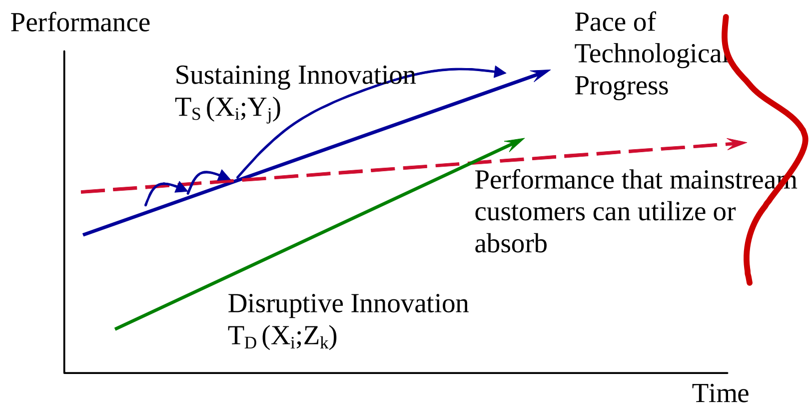
where T_D represents disruptive innovation and T_S represents sustaining innovation

X_i: common product or service attributes associated with both sustaining (T_S) and disruptive

(T_D) innovations; $i = 1,\ldots, n$

Y_j: attributes uniquely associated with sustaining innovation, T_S; j = 1,…, m
Z_k: disruptive attributes; k = 1,…, l

Figure 1. Technological Trajectories of Disruptive and Sustaining Innovations



.

Figure 1 displays the pace of technological progresses or trajectories of disruptive innovation (T_D) and sustaining innovation (T_S). It is the trajectory of the T_D compared with that of the mainstream market demand that is significant (Bower and Christensen, 1995). Utilizing the Internet to facilitate the process of contributing to charities brings with it many new (disruptive) attributes that are not presented in the traditional methods. The performance of those attributes is expected to improve significantly to the point that the mainstream customers will be able to utilize or absorb them. New business models that capitalize on the disruptive attributes and incorporate them into customer value propositions are essential for success. Table 1 compares key performance attributes of the traditional approach of contributing to charity (X_i and Y_j attributes) and the disruptive attributes (Z_k) of e-philanthropy. Some of the most prominent disruptive attributes are, 1) pricing or fee structure - ultimately costing less per transaction, meaning more of the donation will make it to the supported cause or program. 2) Personalization of the process. Through the use of individual accounts, many options can be used to mold the process to the user's personal preferences. 3) Up-to-date news information, financials, and ratings. Along with streamlining the process, moving it online creates the need for processors and charities to make public any and all information about how donations are used, the standing of the organization(s) managing it, and fees taken out of the original donation. 4) Virtually paperless. Funds are transferred electronically, occasionally a service provider may issue a check to a recipient but the majority of the process is electronic, even the donation receipts. 5) Customization. Fundraising campaigns, events and personal accounts are all viable options when a service of this type is implemented. Original setups can allow individuals to easily create customizations needed for specific needs. If not set up originally, new services can be put in place that should take care of needs at little extra cost. 6) Increased efficiency. ETF is common, safe and fast, requiring little human input once a system is in place. Organizations could save on resources by implementing such a system or outsourcing it to a service provider. 7) Ability for smaller organizations to compete for donations. Smaller charities and NPOs can utilize this process and the Internet to inform and educate those interested and compete head to head with larger more recognizable organizations. 8) Knowledge base possibilities allowing NPOs the ability to perform campaigns customized for and targeted at specific markets.

RESPONSES TO STRATEGIC DISRUPTIVE INNOVATION

Many commercial e-philanthropy enterprises were able to capitalize on several disruptive attributes of the Internet and e-philanthropy to mount an “attack” on the traditional NPOs. Even though the technologies underlying the advancement of e-philanthropy are not radical – both computing and communication technologies have been improved incrementally over the past decades, e-philanthropy is disruptive to the traditional NPOs in that it is transforming the rules of competition and inventing new value propositions and business models. Traditional NPOs must respond to the disruptive changes but often find out that they lack the competitive strategy and organizational capabilities to do so. Austin (2001) argues that the benefits of adopting e-philanthropy will depend on how vigorously nonprofit groups embrace the new approach. Charitou and Markides (2003) present a model to assist companies in responding to disruptive innovations. The response depends on two main factors: its motivation and ability to do so. If motivation is low, the response should be to ignore the disruption and focus on the main business. If motivation is high, the appropriate response is dictated by ability and circumstances (e.g., adopt or attack back and disrupt the disruption or embrace the innovation and scale it up).

Mainstream Users of e-Philanthropy

Even with the disruptive performance attributes mentioned above, the e-philanthropy industry in the U.S.A. is still struggling for users. The issue is at what point will mainstream "customers" want to switch to this disruptive strategic innovation. The mainstream is the traditional route of donating money and items to charity. The mainstream is a broad, ambiguous entity that is committed to staying with the status quo. The mainstream as a whole is much too big to define, but it can be measured by three characteristics: 1) number of users; 2) number of transactions; and 3) transaction totals (Blau, 2001a).

Table 1. Comparison of Business Model Performance Attributes

Business Model	Performance Attribute
Common attributes in both traditional and e-philanthropy (online charity/giving) models **(X_i)**	• Trust, Privacy, Security • Reputation • Immediate response to needs of communities across the country • Tax write off • Personal
Attributes and characteristics specific to traditional methods of giving to charity (e.g., by mail, phone, in person) **(Y_j)**	• Established – people know and understand the process • Friends involvement and similar giving tendencies • Proven and reputable • Involvement / Participation – events – tournaments etc. • Paper trail
Characteristics and disruptive attributes specific to the e-Philanthropy model **(Z_k)**	• Fee structure (flat, subscription, etc.) less cost per transaction means more money to cause or program • Personalization of independently managed online accounts and options – empower people • Inform and educate using current news/informationand providing transparency into practices and trends • Virtually paperless EFT – paper receipts and checks by request • Customization of campaigns (targeted and simultaneous), events, accounts and easy

Table 1. Comparison of Business Model Performance Attributes	
Business Model	Performance Attribute
	additions of new services • Increased efficiency using an electronic architecture • Levels competitive environment –organizations can promote themselves side by side with others – broadens reach • Knowledge base – gather, organize, synthesize, and distribute information creating new knowledge and opportunity • Bulletin boards or forums – dialog and feedback • Direct access to "members" - newsletters and email • Capitalize on the virtual value chain

The e-philanthropy industry has a list of unique performance attributes (see Table 1) that have the potential to attract more users. Although the industry can count on three very important factors to increase mainstream users – tragedy, trust, and time, it does not have any control over them. The future for e-philanthropy does not lie with processing online donations or with pursuing a single model of giving (Fix, 2001). Rather, the industry must be able to design various innovative business models that capitalize on the disruptive performance attributes and value propositions to provide real solutions to meet or exceed a customer's expectations.

E-PHILANTHROPY BUSINESS MODELS

A business model is the method of doing business by which a company can generate revenue to sustain itself (Rappa, 2005). It describes the basic framework of a business. It also tells what market segment is being served (who), the service that is being provided (what), and the means by which the service is produced (how) (Chaudhury and Kuilboer, 2002), as well as how the business plans to make money long-term using the Internet (Afuah and Tucci, 2003). Lee and Vonortas (2004) argue that a viable business model in the Digital Economy must follow the fundamental economic principles[13] (i.e., the underlying economic logic that explains how organization can deliver value to customers at an appropriate cost), and also capitalize on the "disruptive attributes" of the Internet and e-commerce.

Customization in the online gift-giving market can be broadly broken down into seven different business models utilized by various NPOs and for-profit companies. Regardless of status, most garner a fee in some form or another for the processing of the funds transfer. Table 2 compares the value propositions and revenue models for each e-philanthropy business model.

Table 2: Business and Revenue Models in e-Philanthropy

Approach/Model	Value Propositions	Revenue Models	Examples
Workplace Giving Intranet (fundraising and transaction processing)	--Free up company resources --Give using credit card, payroll deduction, online check --Trust and security of Employers backing --Automatic periodic deductions – customized campaigns --Volunteer placement opportunities	--Percentage of each donation --Flat Fee (subscription type) --Setup Fees (one time fee)	DonationDepot.com Createhope.com United Way Kindmark.com
One to One Direct to charity Website (transaction processing)	--Money goes directly to specific charity –Easy online form - Secure payment with Credit Card --Setup automatic periodic deductions --Volunteer placement opportunities	--Donations fund operations and programs – Specialty items and collectibles auctions --Grants from government and foundations	Red Cross PETA Various NPOs
One to Many Facilitating Website (transaction processing and search engine)	--Donate to a variety of charities from one website --Setup personal account and automatic deductions --Secure payment with Credit Card --Volunteer placement opportunities	--Foundation and corporate gifts --Fundraising activities --Fee based on percentage of donation --Flat fee	Justgive.org NetworkforGood.org DonationDepot.com

Table 2: Business and Revenue Models in e-Philanthropy			
Approach/Model	Value Propositions	Revenue Models	Examples
Online Community Portal (search engine)	--Reputation – High visibility in time of tragedy --Proven electronic funds transfer system --Personalized web portals --Secure payment with Credit Card --Volunteer placement opportunities	--Financing by corporate foundations --Advertisement --Added value to existing site	N--etworkforGood.org (AOL, Cisco, Yahoo) Amazon / msn.com iVillage.com / Excite
Shopping Online (secondary service)	--Shop online with many big name retailers --Percentage of purchase to charity of choice --Save money with coupons and deals --Secure payment with Credit Card	--Commissions and service Fees --Advertisements --Percentage of purchase price	iGive.com Greatergood.com Shopsthatgive.com
Auction Online (secondary service)	--Changing variety of products in an auction setting. --Sell a personal item with proceeds going to charity --Run an auction campaign for a specific charity --Run auction from your charities own website --Secure payment with Credit Card	--Resale of donated items --Posting fees and percentage of sale amount --Setup Fee – Per event/auction Fee	EBay Goodwill Industries Buynsellit.com Auctionanything.com

Table 2: Business and Revenue Models in e-Philanthropy

Approach/Model	Value Propositions	Revenue Models	Examples
Non Cash Giving (transaction processing)	--Facilitate the movement of gifts in kind --Corporations can unload excess and unused equipment and inventory --Connect donors to charities/NPOs locally and abroad using Internet technology	--Member registration fee --Consulting, and service fees --Processing and shipping --Administration fee --Sales of donated items --NPO – Grants, foundation gifts	Giftsinkind.org Inkindex.com Goodwill Industries

A NEW FRAMEWORK FOR E-PHILANTHROPY

New Philanthropist

It is no longer necessary to have the wealth of a John D. Rockefeller or Bill Gates to make an impact through philanthropy. Individuals, groups and foundations have found a place in the arena once thought reserved for the ultra-wealthy. The philanthropists of today are younger and more involved. They are not waiting until their death to take on a cause or make that big donation. By giving now, rather than transferring their wealth to foundations after death, the new philanthropists can make sure their gifts fund the causes they have chosen in the most efficient way possible (Conlin and Hempel, 2003). Conkey (2006) recognizes that wealthy donors are increasingly opting for a more hands-on approach, giving money on the condition that the charity take their management advice. In some cases, NPOs even agree to let benefactors overhaul their business models, make personnel changes and install financial controls in exchange for new funding. The Michael and Susan Dell Foundation[14] is not looking at supporting established programs but instead wants to create their own, where they can measure the results (McWilliams, 2003). As larger foundations, such as the Dell's or the Bill and Melinda Gates Foundation[15], are moving toward this direct appeal for accountability and results, smaller more specialized charities need to offer this type of transparency to compete for donations.

The new philanthropists follow venture capitalist criteria when it comes to the distribution of funds (Conkey, 2006). This means the organizations that will utilize the charitable funds will have to "produce" before they receive, i.e., they must go through extensive planning and the setting of objectives and goals. They must also be open to input during the project because philanthropists may want direct involvement and even more importantly, they may be able to significantly contribute through knowledge and non-monetary resources. Fund-receiving organizations must also allow the new philanthropist continued involvement after project completion (e.g., a role as consultant). Today's philanthropy is about being involved and insuring that a difference is made. It is for this reason that a framework that defines the new roles and relationships among e-philanthropy industry players is needed.

New Relationships

The Internet and e-commerce have changed the configuration of value chains in almost every industry. As value chains fragment and reconfigure, new opportunities will arise for traditional brick-and-mortar businesses (Evans and Wurster, 1997). Some players within a traditional value chain have changed their roles and value propositions to become electronic intermediaries or infomediaries by utilizing new business models and information technology. Infomediaries convey information on goods and services as well as on companies and industries. Taking advantage of Internet technology capabilities for the collection and processing of information, infomediaries can quickly take on the new role of *innomediaries* (Sawhney, Prandelli, and Verona, 2003), as discussed in the following paragraph.

Olsen, et al. (2001) developed a methodology for comprehensive online donor cultivation and fundraising through building and maintaining one-to-one e-mail relationships with donors and friends of an organization. However, to fully exploit the Internet as an enabler of innovation, companies need to complement their direct channels of customer interaction with indirect, or mediated, interactions. Sawhney, et al. (2003) contend that those points of contact can be carried out by third parties that function as knowledge brokers, helping companies overcome the gaps in knowledge about customers that impede innovation. Innomediaries utilize the concept and practice of the virtual value chain (Rayport and Sviokla, 1995) by aggregating and disseminating customer-generated knowledge, in ways that would help companies improve their innovation processes (Sawhney, Prandelli, and

Verona, 2003). Innomediaries create new knowledge and ideas that can be personally used or sold. The acquisition and utilization of knowledge by one or more organizations could help all do more with less. By applying this concept to the e-philanthropy industry an innomediary could have access to endless relationships and opportunities. This constant flow of new information and then the creation of new knowledge brings with it opportunity.

Two mechanisms that can help e-philanthropy organizations exploit customer knowledge in the service of innovation are: "customer community operator" – an innomediary that specializes in connecting businesses with people who form a community based on common interests (e.g., iVillage, Edmunds, and WebMD), and "innovation marketplace operator" – an innomediary that connects problem solvers or sellers of innovation with seekers or potential buyers (e.g., InnoCentive)[16]. Socially generated explicit and tacit knowledge within communities can be gathered, organized, selected, analyzed and synthesized, and distributed by the innomediary, helping e-philanthropy companies identify and profile influencers and opinion leaders within a customer population, and also adding the potential to speed up the process of design and diffusion of new products and services.

Kintera Inc.,[17] a provider of software as a service that enables NPOs to use the Internet for marketing and revenue generation, currently offers abilities that parallel concepts discussed in this paper. Recently, Kintera purchased a company that possesses a powerful screening tool used by NPOs to find, profile, and rank the "wealth" in their databases. This is a knowledge creation tool. From the outside looking in, Kintera has the ability to be an innomediary or at the very least supply the necessary tools NPOs can utilize to become innomediaries themselves. GreaterGood.com, the United Way and corporate foundations could all utilize an innomediary-based relationship structure, capitalizing on the one item that is the cornerstone of e-philanthropy, knowledge.

Internet Utilization

Internet technology has brought into the home countless new tools and information options. The use of Internet technology on issues of philanthropy is no different. Any individual or group, looking for a cause to support, or in need of supporters, now has a tool for collaboration. The role of an innomediary would be to support, legitimize and increase the functionality of the Internet as a tool, providing aid such as facilitation and virtual workspace. To give parties involved a sense of security through the screening of users and most importantly providing those interested with a place to exchange ideas, information, and expertise, to reach common goals. This will overcome the gaps in knowledge about customers that impede innovation. Philanthropists could “follow the money”, giving them an additional sense of security. Ultimately an *innomediary* structure could be used to build an information and knowledge base that will spawn new innovative business and operations possibilities in philanthropy.

DISCUSSION

This new model of e-philanthropy can provide value added content, resources, and tools that will allow both users and charities to leverage the potential power of the Internet. People are given the ability to act immediately; it is their satisfaction with the process that will dictate the speed at which this industry moves forward. It seems that nonprofits and charities alike may not be maximizing the possibilities of online services and the Internet to stimulate giving. There may be a low awareness of the possibilities that are currently available from service providers and software. Charities and NPOs may also assume they could not safely and securely manage this type of transaction and the handling of information. Users share this concern. Getting the information out and educating the public as to how the system and organizations run are steps all industries moving onto the Internet should take. Sharing personal or financial information online is safe; it is the newness and unfamiliarity that puts potential users on edge.

Online, Internet, intranet, donations or gift giving allows quick and simplified processing of designated funds. With this online process comes the ability to investigate and gather information so the individual user can make a well-informed decision when allocating gifts to charity. The Internet is an important medium for the education of the public on how nonprofits and charities work and what they do. Use of the Internet to compare, contrast and find information on charities is growing. An organization or NPO with a presence on the World Wide Web will enable donors to find, judge and make decisions on where their money goes. Also, charities and NPOs are able to monitor what others are doing, allowing them to keep up to date on campaign strategies and techniques as well as new marketing possibilities. The Internet gives people the option to research nonprofits and their financial activities, thereby making the organizations more accountable to those who finance their operations (e.g., www.guidestar.com).

Charities and NPOs will still need to continue marketing and campaign promotions as they always have. They now have new approaches (e.g. new business models) and tools (e.g. intranet and online communities) to use in soliciting funds from the public. Various e-philanthropy models can be implemented to sustain their business models without the need to rely on major tragedies to raise large amounts of donations.

CONCLUSION

The e-philanthropy revolution is here to stay, and it will transform charitable giving in as profound a way as technology is changing the commercial world (Austin, 2001). The facilitation of donations to NPOs and charities is an old market with new possibilities. e-philanthropy is a disruptive strategic innovation that has fundamentally changed the competition in the traditional philanthropic industry. This innovation will eventually overtake the traditional gift-giving market. The question is which new e-philanthropy business model or combination of models will come out on top.

e-philanthropy comes in a variety of customizable tools that can be used individually or pooled with other innovative and traditional methods to fulfill specific needs. It allows a cost effective process to conveniently move money to a desired recipient, allowing more money to go to programs where it can do the most good. The benefits to the gift recipients should be larger portions of donations retained for the specified use, as well as an inexpensive route for them to solicit and receive money directly from individuals.

There is a need for charities, NPOs, and organizations to look at new types of relationships with benefactors. The new philanthropist wants to be involved and emerge with self-gratification on a job well done. These relationships need to be cultivated and built into long-term partnerships, not just one-offs satisfying someone's individual desire to do well. Finally, to operationalize the concept of disruptive innovation, this paper proposes a new method that can be applied to assist innovation managers and entrepreneurs in identifying the unique attributes and designing an innovation business model in order to capture the full benefits of a disruptive innovation. In addition, this paper also proposes a new system that utilizes the concept of the virtual value chain and innomediation to produce new knowledge, services or outlets for users to advance their needs.

REFERENCES

Afuah, A. & C.L Tucci (2003). *Internet business models and strategies: Text and cases*. (Second Edition). New York, NY: McGraw-Hill/Irwin.

Arthur, W.B. (1996). Increasing returns and the new world of business. *Harvard Business Review*, 74(4), 100-109.

Austin, J.E. (2001). The e-philanthropy revolution is here to stay. *Chronicle of Philanthropy*, 13(10), 72-73.

Blau, A. (2001a). Internet giving: Not the perfect revolution. *Chronicle of Philanthropy*, 14(2), 85-86.

Blau, A. (2001b). More than bit players: How information technology will change the ways nonprofits and foundations work and thrive in the information age. *A Report to the Surdna Foundation*, May 2001. Retrieved September 29, 2003, from http://www.surdna.org/documents/morefinal.pdf

Bower, J. L. & C.M. Christensen (1995). Disruptive technologies: Catching the wave. *Harvard Business Review*, 73(3), 43-53.

Charitou, C.D. & C.C. Markides (2003). Responses to disruptive strategic innovation. *Sloan Management Review*, 44(2), 55-63.

Chaudhury, A, & J. Kuilboer (2002). *E-Business and e-commerce infrastructure: Technologies supporting the e-business initiative*. New York, NY: McGraw-Hill/Irwin.

Christensen, C.M. (1997). *Innovator's Dilemma: When Technologies Cause Great Firms to Fail*, Boston, MA: Harvard Business School Press.

Christensen, C.M. & M. Overdorf (2000). Meeting the challenge of disruptive change. *Harvard Business Review*, 78(2), 67-76.

Conkey, C. (2006). Strings attached: Along with their big bucks, rich donors want to give charities their two cents. *Wall Street Journal*, July 3, B1.

Conlin, M. & J. Hempel (2003). The top givers: Today's philanthropists aren't leaving the good works to future generations – they're making their mark now. *BusinessWeek*, December 1, 78-84.

Evans, P.B. & T.S. Wurster (1997). Strategy and the new economics of information. *Harvard Business Review*, 75(5), 71-82.

Fix, J.L. (2001). Non-virtual reality hits giving sites. *Chronicle of Philanthropy*, 13(17), 9-11.

Hruby, L., D.E. Blum, & M. Voelz (2001). All aboard. *Chronicle of Philanthropy,* 13(17), 8-10.

Lee, C.-S. (2001). An analytical framework for evaluating e-commerce business models and strategies. *Internet Research*, 11(4), 349-359.

Lee, C.-S. & N.S. Vonortas (2004). Business Model Innovation in the Digital Economy. In Georgios Doukidis, Nikolaos Mylonopoulos, and Nancy Pouloudi (eds.), *Social and Economic Transformation in the Digital Era*, Hershey, PA: Idea Group Publishing, 164-181.Markides, C.C. (1997). Strategic innovation. *Sloan Management Review*, 38(3), 9-23.

McWilliams, G. (2003). Dells plan hands-on approach to charity donations, programs. *The Wall Street Journal*, New York N.Y., October 2, B10.

Miller, W. L. & L. Morris (1999). *Fourth Generation R&D: Managing Knowledge, Technology, and Innovation*, New York, NY: John Wiley & Sons, Inc.

Olsen, M., M.L. Keevers, J. Paul & S. Covington (2001). E-relationship development strategy for the nonprofit fundraising professional. *International Journal of Nonprofit and Voluntary Sector Marketing*, 6(4), 364-373.

Porter, M.E. (2002). The competitive advantage of corporate philanthropy. *Harvard Business Review*, 80(12), 57-68.

Rappa, M. (2005). Business models on the Web. *Managing The Digital Enterprise*. Retreated February 27, 2005, from http://digitalenterprise.org/models/models.html.

Rayport, J.F. & J.J. Sviokla (1995). Exploiting the virtual value chain. *Harvard Business Review*, 73(6), 75-85.

Sawhney, M., E. Prandelli & G. Verona (2003). The power of innomediation. *Sloan Management Review*, 44(2), 77-82.

Sinha, I. (2000). Cost transparency: The Net's real threat to prices and brands. *Harvard Business Review*, 78 (2), 43-50.

Tapscott, D. (1996). *The Digital Economy: Promise and Peril in the Age of Networked Intelligence*. New York, NY: McGraw-Hill.

Wallace, N. (2002). Outlook for online donations is cloudy, experts say. *Chronicle of Philanthropy*, 14(11), 27.

ACKNOWLEDGMENTS

This research was supported by the *ePLU E-Commerce & Technology Management Center* (http://eplu.plu.edu) in the School of Business at Pacific Lutheran University. A preliminary version of this article was presented in October 2006 at the Allied Academies International Conference in Reno, NV. The authors wish to thank the anonymous reviewers and the conference participants for their valuable comments and suggestions.

ENDNOTES

[1] Giving USA 2005, the Annual Report on Philanthropy and charitablechoices.com.

[2] "Giving and volunteering in the United States: Findings from a national survey, 1999", Independent Sector, Washington D.C. [Online]. Available at http://www.independentsector.org/GandV/default.htm

[3] "Giving and Volunteering in the U.S. 2001: A survey by Independent Sector", Washington D.C. [Online]. Available at http://www.independentsector.org Performed by The Independent Sector and conducted in May and June of 2001 with 4216 people, all at least 21 years of age.

[4] "Re: Development", *Fund Raising Management*, 32(3), 2001.

[5] Ibid.

[6] Such consolidations already have been occurring. For example, GreaterGood.com acquired the Hunger Site in 2000. In addition, GreaterGood.com also consolidated the breast cancer site, the rainforest site, the animal rescue site, and the child health site into a single portal site for e-philanthropy.

[7] *Prosumption* is the term to describe the convergence of design with development process and the production of goods and services by customers in the e-commerce environment.

[8] Digital assets are information about customers (e.g., purchasing patterns and profiles). A firm that exploits the Internet should build and utilize its digital assets in order to provide customer value across many different and disparate markets.

[9] The advance in network and storage technologies gives customers the feeling that it has infinite virtual capacity to serve them.

[10] In the digital economy, the traditional trade-off between richness and reach in information exchange no longer exists. Information can reach many customers or business ecosystem partners through the Internet without sacrificing the richness of the contents

[11] The vast amount of information about prices, competitors, and features that is readily available on the Internet helps buyers “see through” the costs of products and services.

[12] Value generated in Internet-enabled business transcends traditional industrial sectors.

[13] See Lee and Vonortas (2004) for detailed discussion of the fundamental economic principles in the Digital Economy (e.g., new economies of scale and scope, switching costs, transaction costs, and pricing and revenue models).

[14] http://www.msdf.org

[15] http://www.gatesfoundation.org

[16] Sawhney, Prandelli, and Verona (2003) provide detailed discussion and several examples of these two types of innomediaries.

[17] http://www.kintera.com

ARTICLES for Volume 5, Number 2

EMPIRICAL EVIDENCE ON EBAY BIDDING STRATEGIES

Michael Hergert, San Diego State University

ABSTRACT

EBay has become a global force in electronic commerce. Numerous books and articles have offered advice to potential buyers and sellers as to the most effective strategies for improving their prices. This study analyzes price data from eBay transactions to measure the impact of various buyer and seller strategies on ultimate price. The measurement of seller trustworthiness is studied and its role in price determination is analyzed. It appears that some aspects of the conventional wisdom about eBay tactics are upheld while others do not seem statistically significant. With the growing use of on-line auctions and electronic markets, these factors are likely to grow in importance in the future.

INTRODUCTION

In just a few years, eBay has emerged as a dominant force in e-commerce. Founded in September 1995, eBay has become the world's largest online marketplace. The eBay community includes more than a hundred million registered members from around the world. People spend more time on eBay than any other online site, making it the most popular shopping destination on the Internet. As stated in their Annual Report, eBay's mission is to provide a global trading platform where practically anyone can trade practically anything. On an average day, there are millions of items listed on eBay. The annual revenues for eBay are greater than the gross domestic product of more than half the nations in the world (Hof 2003).

EBay has become a global phenomenon. EBay has local sites that serve Australia, Austria, Belgium, Canada, China, France, Germany, Hong Kong, India, Ireland, Italy, Malaysia, the Netherlands, New Zealand, Poland, the Philippines, Singapore, South Korea, Spain, Sweden, Switzerland, Taiwan, the United Kingdom, and the United States. Furthermore, eBay is a major business success. EBay has a market capitalization of over $58 billion, giving it a valuation greater than General Motors and Ford put together. If the 430,000 individuals who earn all or most of their income selling on the site were eBay employees, it would be the second largest employer in the Fortune 500 list, behind only Wal-Mart (Sellers 2004).

Numerous books and articles have suggested various bidding and selling strategies to improve the transaction prices of desired products. This study analyzes a data set of eBay transactions to determine whether or not these proposed strategies result in better prices. In particular, the effects of sniping (last minute bidding), opening bid manipulation, seller track record, handling/shipping costs substitution, and other commonly observed approaches are analyzed. Prior research in this area has been ambiguous. Most studies have focused on a particular aspect of transactions and have not controlled for other variables. This study develops a multivariate model of auction price determination. There is evidence to suggest that proper use of specific strategies can produce superior net results for both buyers and sellers.

BACKGROUND

The rise of electronic commerce has led to numerous online markets where buyers and sellers have never met. Much of the previous research into online auctions has focused on the role of trust. Trust is an important factor in every marketplace, but even more so in electronic commerce (Ba et al. 1999, Brynjolfsson and Smith 2000). The impersonal and anonymous nature of electronic commerce creates a fertile environment for the manipulation of transactions and potential fraud. Almost all online auction sites address this issue through the use of information exchange about the background and experience of buyers and sellers. The most notable of these systems is EBay's Feedback Forum. Online feedback mechanisms allow buyers and sellers to report their satisfaction with each transaction in a public forum. These mechanisms can serve to build trust among the participants in electronic commerce (Walden 2000, Ba and Pavlou 2002). The fundamental problem is the information asymmetry between buyers and sellers (Akerlof 1970). When bidders view a product at an auction site, they do not have the opportunity to inspect the product and directly observe the quality (Fung and Lee 1999). This unequal access to information can lead to fraud or market failure. Buyers are exposed to additional risk because they must rely on incomplete or potentially distorted information provided by the seller (Lee 1998). The building of trust is most crucial when two factors are present: risk and incomplete/asymmetric information (Swan and Nolan 1985). Both of these factors are more prevalent in electronic commerce than traditional markets. Online feedback mechanisms, also known as reputation systems (Resnick et al. 2000), are using the Internet's bidirectional communication capability to artificially engineer large-scale, word-of-mouth networks in which individuals share opinions and experiences (Dellarocas 2003). Despite the wide use of reputational mechanisms such as EBay's Feedback Forum to promote trust, there has been little empirical evidence as to whether these mechanisms actually induce trust or create any desirable outcomes such as increased bidding or higher final prices (Ba and Pavlou 2002). This study will investigate several measures of trustworthiness and analyze the effect on transaction prices.

The sheer magnitude of the data required for assessing reputations is daunting. At any given moment, eBay will list several million products for sale to over 100 million potential customers. An ongoing trading relationship between buyers and sellers is one way to develop trust (Ba and Pavlou 2002), but the vast majority of eBay transactions are one shot deals. Resnick and Zeckhauser (2002) found that 89% of all buyer-seller pairs conducted just one transaction during the 5 month period under investigation.

This study will analyze the impact of buyer and seller reputation on final transaction prices. Earlier studies have tried to measure the importance of reputation in determining both the probability of sale and the ultimate price in online transactions. The results have been mixed. Most studies show at least a weak positive impact of reputation on price, but some have found the opposite effect (Dellarocas 2003). Part of the problem in measuring the impact of reputation is the need to control for other variables. In this study, reputation will be measured in several ways and other factors affecting price will be incorporated into a multivariate model.

HYPOTHESES AND DATA SET

To study the determinants of prices in eBay transactions, the first step was to define a product class. Prior studies have generally chosen a single product in order to hold constant for the effect of product attributes. Products sold on eBay can vary between disposable items selling for less than one dollar to real estate selling in the tens of millions. The price level plays a significant role in determining the importance of reputation. The impact of feedback profiles on prices has been found to be more important for riskier transactions and more expensive products (Dellarocas 2003, Lee et al. 2000). The product chosen for this study was the DVD of the first season of The Sopranos (HBO, 1999). This product was chosen for several reasons. First, it is readily available from a variety of online and traditional retailers. It is an essentially standardized product, but may still raise quality issues in the mind of the buyer. Approximately half of the transactions studied were for used DVDs, thus raising the possibility of purchasing a damaged disc. Many of the sellers were located abroad. This could create suspicion that the DVDs could be bootlegs or be in a format that is incompatible with North American DVD players. The average price was approximately $28 (including used versions). This is high enough to be treated somewhat seriously by potential buyers, but low enough to not require extensive shopping or inspection. An older DVD (vintage 1999) was chosen to prevent any distortions due to time effects from recent releases.

The following variables and hypotheses were studied:

PRICE. Prices were collected for 147 transactions. The average price for new versions of the DVD was approximately $32. This is considerably below the retail price ($59.98) and the price charged by traditional outlets (Wal-Mart $44.88, Barnes & Noble $43.18). Other online vendors were also more expensive. Amazon.com and Target.com both charged $44.99. The eBay selling price served as the dependent variable in this study.

Reputation. As discussed previously, reputation is crucial in online markets. EBay has recognized this fact and developed several mechanisms for measuring seller trustworthiness. This study used four measures:

(1) SELLER SCORE. At the conclusion of each transaction, buyers are given the opportunity to rate their satisfaction with the seller as (-) negative, (0) neutral, or (+) positive. Buyers provide a rating for approximately 52% of all transactions (Resnick and Zeckhauser 2002). EBay then calculates a Seller Score as the number of positive transactions with unique buyers that a seller has completed. This score measures both the experience and reliability of a seller. In this study, the seller scores ranged between 0 and 9,665. A score of zero means that this was the first eBay sale for that seller. A score of over 9,000 is indicative of a commercial enterprise selling to many different buyers who were satisfied with their experience. The hypothesis in this study is that a high seller score is likely to lead to higher final transaction prices. All other things equal, buyers should be willing to pay a price premium to do business with individuals (or companies) with an established positive track record. This is consistent with the findings of Dewan and Hsu (2002) and Melnik and Alm (2002) among others.

(2) SELLER % POSITIVE. In addition to calculating a Seller Score, eBay also provides information on the total number and distribution of ratings for

each seller. From this, it is possible to calculate the percentage of positive ratings that each seller has received. Unlike the seller score, this variable does not depend on the volume of activity, but simply measures the reliability of the seller. This variable should play a similar role to Seller Score, but raises the possibility of overstating the trustworthiness of a seller who has not engaged in many transactions. The hypothesis is that a high percentage of positive ratings will be associated with higher transactions prices.

(3) YEAR OF FIRST TRANSACTION. It is relatively simple to establish an eBay account for the purposes of buying or selling. As a result, there is the constant threat of dealing with disreputable individuals who simply create aliases for fraudulent purposes and disappear into cyberspace after failing to perform. As a check on this, eBay provides background information on buyers and sellers including the year of their first recorded transaction. A long track record of commerce will add credibility to a seller's reputation. In the data set used in this study, seller's ranged between 0 and 10 years of experience. The hypothesis is that buyers will be willing to pay a price premium to deal with well-established sellers having long track records. Conversely, buyers will expect a price discount from sellers who have no established record on eBay.

(4) NUMBER OF TRANSACTIONS. In addition to the length of time that a seller has participated on eBay, buyers may also prefer to deal with sellers who have engaged in many transactions. EBay provides background information on the cumulative number of deals that a seller has participated in. The hypothesis for this variable is similar to the other three measures of seller trustworthiness. Buyers will be willing to pay a price premium to deal with sellers who have an extensive track record on eBay. Conversely, they will expect a price discount if they are dealing with an unknown quantity.

NUMBER OF BIDS. The conventional wisdom on eBay is that sellers should make every effort to attract attention to their auctions and create a critical mass of bidders. As will be discussed later, there are a variety of ways to attempt to do this. The objective is to ignite a bidding war that encourages bidders to offer the highest prices they are willing to pay. Data was collected on the total number of bids for each of the 147 auctions in this study. This number varied between 1 and 28. The hypothesis is that auctions with a larger number of bids have succeeded in attracting the attention of many possible buyers and will lead to a higher final price. However, there is a counter-strategy from buyers that could mitigate this effect. The most common pattern for bidding on eBay is to have a relatively low starting bid that is followed by gradually increased prices up until the final moments of the auction. The last few minutes is typically when the real action starts. The bids generally escalate rapidly until the final few seconds. Some buyers will try to derail this process by making a somewhat high preemptive bid in the early period of the auction. This can discourage other potential bidders from even considering getting involved as they might assume the product is getting expensive early on. The bargain hunters will move on to other auctions for similar products and opt out of becoming participants in a bidding war.

SHIPPING COST. After winning an auction, the buyer must pay for the product to be shipped. The seller can set any shipping cost desired. In some cases, free shipping is offered as an enticement to attract attention. In other cases, sellers will treat the shipping as a profit center and charge an amount well in excess of the actual cost. For example, the shipping costs charged for the auctions in this data set varied between 0 and $15.27 (with a mean of $5.98). In an efficient market, neither of these seller strategies should have much impact. A rational buyer should only care about the total price paid (sales price plus shipping cost) and would be indifferent between the split between them. Previous research has suggested that buyers might be fooled by this tactic (Hossain and Morgan, 2006). A low sales price might encourage buyers to feel they have found a bargain, even if the savings is immediately spent on higher shipping. The hypothesis in this study is that buyers are rational and will be indifferent to artificially high or low shipping costs.

NEW. This study analyzed both new DVDs and resales. As such, it was important to control for the condition of the product. A dummy variable of (0,1) was used to distinguish between new and used discs. About half of the transactions were for new DVDs. The coefficient on this variable will indicate the price premium for new discs over used ones.

FOREIGN. About 15% of the transactions in this study originated from abroad. This raises another dimension of trust: whether or not the DVD is a legitimate copy. Some of the foreign DVD sellers on eBay are notorious bootleggers. Many buyers will shun these auctions in principle while others may be willing to buy from abroad if the price is right. The foreign discs also have the reputation of having format problems. DVD formats vary around the world. Many of the foreign sellers are located in China and while they may describe their products as compatible with North American DVD players, it isn't always the case. The hypothesis in this study is that buyers will expect a price discount for DVDs that originate outside of the United States.

BUYER EXPERIENCE. Just as the seller track record can affect price, so can the experience of the buyer. Sophisticated buyers are more likely to employ tactics that lead to lower final prices, such as making last minute bids (sniping) or doing extensive comparison shopping. EBay provides bidder background information including the number of transactions each buyer has completed in the past. This variable was used as a measure of buyer experience and sophistication. In this data set, buyer experience varied between 0 and 806 transactions, with a mean of 83. The hypothesis is that auctions won by experienced buyers are more likely to have lower final prices.

LENGTH OF AUCTION. As part of establishing an auction, the seller can choose how many days to list the product. The conventional wisdom is that auctions must be long enough to allow as many potential buyers as possible to browse across your item. However, there is a risk that impulse buyers will shun auctions that will not close for many days. Since much of the bidding activity takes place in the final hours of an auction, it is not clear that additional days of listings will contribute much to the final price. In this study, the length of auctions varied between 1 and 10 days. The hypothesis is that longer auctions will provide a greater opportunity to attract buyers and will result in higher final prices.

STARTING BID. Sellers are also given the opportunity to specify a starting bid. On their web site, eBay encourages sellers to set a low starting bid as a way to attract attention. They warn that asking for a high starting bid can choke off interest in the product and lead to lower final prices. However, it is not clear that buyers who are attracted to artificially low prices are likely to end up bidding at higher levels. These bargain hunters may simply fade away as soon as prices approach competitive levels. On the other hand, once a specific auction has entered the evoked set of possibilities that buyers are actively monitoring, there is the possibility that they will become vested in the transaction and contribute to a bidding war. In this data set, the starting bids set by sellers varied between $.01 and $40. The hypothesis is that a low starting bid will attract the attention of more potential buyers and ultimately lead to a higher final price.

SNIPE. Perhaps no other phenomenon on eBay has received as much attention as "sniping"- the practice of waiting until the very last seconds of an auction and bidding slightly above the going price to steal away an item. There are even specialized software programs and web sites to automatically do this for buyers. Such behavior can incur the wrath of other bidders, but does it lead to better prices? In this study, a dummy variable was assigned a value of 0 or 1 depending on whether or not the winning bid was submitted in the final two minutes of the auction. Approximately 37% of the auctions were "sniped" in this data set. The hypothesis is that an auction that received a winning bid during the final two minutes will have a lower final price.

PICTURES. EBay encourages sellers to do everything they can to attract attention. They offer additional promotional opportunities, such as including photos of the product as part of the listing. In this study, the product is relatively standardized, so the inclusion of photos might not have a strong effect. However, anything which encourages more buyers to visit or monitor a particular auction can contribute to a higher closing price. Prior studies have shown a relationship between the inclusion of auction pictures and the creation of a herd instinct that can result in higher prices (Stern and Stafford, 2006). The hypothesis is that auctions that included photos of the product are likely to exhibit higher final prices.

RESULTS

The preceding variables were analyzed using multivariate regression. Some typical results are shown in Table 1. Approximately 43% of the variation in the dependant variable (PRICE) can be explained by the six included independent variables (NUMBER OF BIDS, SHIP COST, NEW, STARTING BID, SNIPE, SELLER % POSITIVE). All of these variables were significant at the 95% confidence level. Most of the findings are consistent with the hypotheses described previously.

The results from the variables relating to seller trustworthiness are of particular interest. Four different measures of trust were used (SELLER % POSITIVE, SELLER SCORE, START YEAR, # OF TRANSACTIONS). Of these, the percentage of positive ratings received by sellers was by far the most significant. Each of these measures addresses a slightly different aspect of trust. Of the four, SELLER % POSITIVE is the only one that is sensitive to the proportion of negative ratings that a seller has received. The other variables are more dependent on volume and duration of seller experience. This means that sellers must be very careful not to alienate buyers during their dealings. Any blemishes on their record that lower their percentage of positive scores will lead to lower selling prices.

The importance of attracting a large number of bids was also demonstrated. There was a strong relationship between the number of bids and the final price. This supports the hypothesis that sellers need to attract the attention of buyers and to encourage them to submit bids, even if they are low. The herding behavior identified in other studies seems to play an important role in price determination. This may also explain the findings on starting bids. It appears to be helpful to have a lower starting bid as a means of attracting attention to the auction.

Table 1: Determinants of Auction Prices

Dependent Variable: PRICE

Independent Variables	Coefficient	t Statistic	Significance
CONSTANT	14.460	3.57	.001
NUMBER OF BIDS	.836	7.31	.000
SHIP COST	-.937	4.59	.000
NEW	2.198	2.34	.021
STARTING BID	-.442	6.71	.000
SNIPE	-2.220	2.30	.023
SELLER % POSITIVE	.068	1.97	.050

Adjusted R Squared: .431

The findings on shipping cost are somewhat surprising. If buyers are rational, there should be a one-to-one discount between higher shipping costs and lower selling prices. This isn't quite the case. The coefficient on shipping costs is -.937, indicating that an additional $1.00 of shipping cost will only reduce the sales price by 94 cents. In other words, the seller would be wise to offer products at artificially low selling prices and to recoup the difference in inflated shipping costs. There is approximately a 6% profit gain from shifting the revenue from sales price to shipping cost. This may explain why it is common to see sellers employing this tactic on eBay, despite the readily observable nature of this strategy.

Not surprisingly, new DVDs sell for more than used ones. However, the difference is not very big. On average, new DVDs only command a $2.20 premium over used ones. This might be due to the fact that the new DVDs sold at such low prices on eBay relative to retail and traditional outlets. The average sales price of the new DVDs in this study was 47% below the retail price and 29% below the price charged by bricks and mortar outlets.

It also appears that sniping does work. In fact, it creates a significant discount for the buyer when used. The auctions which had a winning bid during the final two minutes would appear to provide a $2.22 reduction in the final price. This is a very significant percentage of the entire product price. Buyers are well advised to control their impulses to bid until the very end of an auction. Patience pays off.

There were several non-findings in this study. Some of the variables tested did not prove significant. Surprisingly, there was no discount for foreign DVDs when shipping costs are controlled for. Buyer experience did not play a direct role, although it may have been picked up by the sniping variable. Several of the reputation variables seemed weakly related to price. In general, it appeared that buyers were more sensitive to reliability (as measured by percentage of positive ratings) than to seller experience or track record. Length of auction and inclusion of pictures were not significant.

CONCLUSION

EBay has become a major force in e-commerce. They have developed a brilliant business model that has become part of the lives of millions of consumers. This study has analyzed the factors which determine final transaction prices in the eBay marketplace. In general, many elements of the conventional wisdom were validated. There are specific tactics that sellers can employ to improve the prices they receive. Likewise, buyers can affect final prices through the use of sniping. There are several ways in which future research could extend the results of this study. A compelling theoretical model of the online bidding process has still not been developed. This study controlled for product attributes by examining a single product. Future cross-sectional studies will enhance our understanding of the interplay between product complexity and the bidding process. The role of trust and reputation needs further investigation. The emergence of online markets has created a paradigm shift in how buyers and sellers assess each others attractiveness as partners. EBay's Feedback Forum was an innovative first step down this road, but there is certainly room for better tools to improve the efficiency of online commerce. There are now over 1,600 online auction sites (Brint, 2003) that are seeking better mechanisms to overcome buyer and seller reluctance. Future research will improve our understanding of the dynamics of online marketplaces.

REFERENCES

Akerlof G. (1970). The Market for Lemons: Quality Under Uncertainty and the Market Mechanism. *Quarterly Journal of Economics,* (84), 488-500.

Ba, S., Whinston, A. B. & H. Zhang (1999). Building Trust in the Electronic Market Using an Economic Incentive Mechanism. *Proceedings of the 1999 International Conference on Information Systems*, Charlotte, NC.

Ba, S. & M. Pavlou (2002). Evidence on the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly* (26:3), 243-268.

Brint, A. (2003). Investigating Buyer and Seller Strategies in Online Auctions. *Journal of the Operational Research Society*, (54), 1177-1188.

Brynjolfsson, E. & M. Smith (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers, *Management Science* (46:4), 563-585.

Dellarocas, C. (2003). The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science* (49:10), 1407-1424.

Dewan, S. & V. Hsu (2002). Adverse Selection in Reputation-Based Electronic Markets: Evidence from Online Stamp Auctions. Working Paper, University of California, Irvine.

Fung R. & M. Lee (1999). EC-Trust: Exploring the Antecedent Factors. *Proceedings of the Fifth Americas Conference on Information Systems*, Milwaukee, WI, 517-519.

Hof, R. D.(August 25, 2003). The eBay Economy. *Business Week* , 125-128.

Hossain, T. & J. Morgan (2006). *Advances in Economic Analysis and Policy*.

Lee H. G. (1998). Do Electronic Marketplaces Lower the Price of Goods? *Communications of the ACM* (41:1) 73-80.

Lee, Z., Im, I & J. Lee (2000). The Effect of Negative Buyer Feedback on Prices in Internet Auction Markets. *Proceedings of the ICIS*, 286-287.

Melnik, M. I. & J. Alm (2002). Does a Seller's Reputation Matter? Evidence from EBay Auctions. *Journal of Industrial Economics*, (50:3) 337-349.

Resnick, P., Zeckhauser R., Friedman, E. & K. Kuwabara, (2000). Reputation Systems. *Communications of the ACM*, (43:12) 45-48.

Resnick, P. & R. Zeckhauser (2002). Trust Among Strangers in Internet Transactions: Empirical Analysis of EBay's Reputation System. *The Economics of the Internet and E-Commerce*. *Advances in Applied Microeconomics*, (11) JAI Press.

Sellers, P. (May, 2004). EBay's Secret, *Fortune*, (150:8) 160-1267.

Stern, B. & M. Stafford (2006). Individual and Social Determinants of Winning Bids in Online Auctions. *Journal of Consumer Behaviour*, (5:1), 43-56.

Swan, J. E. & J. Nolan (1985). Gaining Customer Trust: A Conceptual Guide for the Salesperson. *Journal of Personal Selling and Sales Management* (5), 39-48.

Walden, E. (2000). Some Value Propositions of Online Communities. *Electronic Markets*, 10, 4.

SELLING ON EBAY: PERSUASIVE COMMUNICATION ADVICE BASED ON ANALYSIS OF AUCTION ITEM DESCRIPTIONS

Claudia Rawlins, California State University, Chico
Pamela Johnson, California State University, Chico

ABSTRACT

The sellers of the over twelve million products listed on eBay every day are hoping to receive the highest possible price. An analysis of product descriptions leads to the conclusion that readability is the most important contributor to an active sale. Clear description of product and product quality should be addressed in complete sentences. Buyers prefer a product description where the seller sounds knowledgeable but does not use abbreviations or jargon.

INTRODUCTION

> *"There is nothing in the world that some man cannot make a little worse and sell a little cheaper, and he who considers price only is that man's lawful prey."*
>
> —John Ruskin (1819 – 1900)

Although many companies offer auctions on the Internet, eBay is by far the largest. Founded in 1995, the company has grown rapidly and profitably. Total consolidated net revenue for 2002 was $1.21 billion, a 62% increase over 2001. On any given day, there are more than 12 million items listed in 18,000 categories, and more than 42 million buyers and sellers are active eBay members. In 2002, eBay's members transacted $14.87 billion in gross merchandise sales (eBay.com).

Every seller is hoping to get the highest possible price for his or her product. The auction format is the classic "free market economy" where all buyers have access to all sellers. In addition, eBay saves all auction results for several weeks, so buyers and sellers have access to all information about recent completed sales. As a result, according to classic economic theory, price should emerge as the perfect balance between supply and demand.

If eBay provides the perfect marketplace, then all sellers should receive similar prices for similar products. But they do not. Some sellers receive no bids at all. Some auctions generate active bidding and higher ending prices. This paper attempts to answer, in part, the question: Can the seller do anything to increase the price the buyer is willing to pay?

BACKGROUND

Many authors provide eBay sellers with advice on how to use eBay's powerful website tools. They also offer suggestions for how to promote products. Sinclair (2001, pp. 258-264) identifies the following as being important to a successful auction item description:

Specifically identify the item being auctioned.

1. State the condition of the item.
2. Provide one or two major benefits of the item to potential bidders.
3. Provide information about the product.
4. Provide information about purchasing (payment methods, shipping information, etc.)
5. Promote yourself and your other auctions.
6. Ask for a bid.

Collier, Woerner, & Becker (2002, pp. 200-201) suggest an item description should do the following:

1. Accentuate the positive.
2. Include the negative.
3. Be precise about all the logistical details of the post-auction transaction.
4. Promote your other auctions.
5. While you're at it, promote yourself too.
6. Wish your bidders well.

Collier (2002, pp. 208-218) stresses the importance of an excellent quality photograph for reaching the largest number of potential bidders and thus increasing the final bid price. The eBay website provides answers to hundreds of Frequently Asked Questions about posting auctions, and visitors to eBay chat rooms share advice for getting the best price. For example, correct spelling is stressed so that buyers can find items using eBay's search engine. In general, the advice is the advice given to all writers of persuasion, and is based on an intuitive sense of what works. However, none of these sources provide research-based information. Based on actual auctions, what are the critical factors in increasing selling price? That is the issue this research attempts to address.

METHODOLOGY

Over the period of January 14 – February 18, print outs of completed auctions for "Roy Rogers original 1948 – 1951 Dell" comic books were collected. Auctions were then matched for the following characteristics:

- number of days the auction was available (either 7 or 10 days)
- exact issue number
- same condition

A total of 168 auctions were considered, with from two to five auctions in a matched set. Fifty-nine different sets were analyzed. For each auction, the following data were collected:

- auction ending date
- number of days auction was available
- issue number
- condition of issue
- photograph (yes/no) and quality of picture
- start price
- end price
- number of bids
- number of hits (when available)
- analysis of the product description (see following)
- analysis of layout (see following)

The product descriptions were analyzed using a simplified discourse analysis methodology (see Berkenkotter & Huckin, 1995; Powers, 2001; Stillar, 1998; Stoddard, 1991 for discussions on elements of discourse analysis.) The number of each of the following was determined: total words, sentence fragments, complete sentences, personal pronouns (self directed/other directed), words focused on product description/self promotion/shipping, handling & paying/goodwill, grammar errors, spelling errors, abbreviations used, genre-specific jargon, exclamation points. The overall tone was determined (positive/neutral/negative).

In addition, elements of visual style were noted. Such elements included whether the description was left-justified/centered/right-justified, font size, font type, black and white or color, one column or two, all capitals.

Comparisons of all these elements were made between the auction in each matched set with the highest ending price and the auctions with lower ending prices.

CONCLUSIONS

Based on the comparisons of actual auction prices, there is much a seller can do to influence bidders to bid more. These are discussed in order of significance.

- First, and critical, is a good quality picture. Auctions without pictures end with no bids, in the case of Roy Rogers comics, 94 percent of the time. When the seller included a picture, sharper and more colorful pictures were associated with the higher ending price 71 percent of the time. The brighter and more in-focus the picture, the better. Using a flash almost always washes out a portion of the picture. Natural light or studio light makes much better pictures.
- Second, and essential, is readability. The eBay buyer has little patience with anything that makes reading the product description difficult. The most problematic include small font size (12 point is best in all situations), centered text (which is unfortunately the eBay default), capital letters except for the product name, misspelled words, lack of paragraphing, jargon and abbreviations. Readability problems resulted in no bid 33 percent of the time and in lower bids 88 percent of the time.

  As a result, the best product description will use 12-point type, will use all capital letters only for the product name, will be left-justified, and will separate the description into at least three paragraphs.

Other factors which result in poor readability but were not common enough to be analyzed included type so large that the reader had to scroll down to see it all, animation and/or music which slowed loading time, poor typing skills so that spacing was wrong (i.e. "VoL.2,#3"), and the excessive use of exclamation points.

- The best number of words focused on product description was 40-55. In this study, the fewest number of words describing the product and its quality was 9 and the most was 130.
- The best number of words focused on shipping, handling, and paying was 11-20. In this group of comic auctions, the fewest was 0 and the most was 169.
- The seller promoting his or her other auctions made no significant difference in sales price.
- Goodwill (such as "thank you" or "good luck with your bidding") or personal pronouns (such as "you" and "your") made no significant difference in sales price. It may make a difference in the feedback the seller receives—and thus ultimately in the number of buyers who are willing to work with that seller--but that wasn't considered in this study.
- Complete sentences were more effective than sentence fragments 57 percent of the time.
- The eBay bidder prefers that the seller have good product knowledge while using no abbreviations or jargon. Sellers who explain that they know little about the general product type they are selling, using words like "probably" or "might" or "I'm not an expert," received the lower price 89 percent of the time. At the same time, bidders do not respond well to abbreviations like "COA" or "Condition 3." Sellers who use such shorthand received the lower price 73 percent of the time.
- So few sellers develop good buyer benefits that the numbers were too small to be analyzed. Of the 168 auctions included in this study, only four sellers included benefits such as "The colors on this comic are so bright it would make beautiful wall art if framed." The lack of product descriptions that included benefits may provide an opportunity for sellers to develop competitive advantage.

- ♦ In only one case was the seller's screen identification name a possible problem in his or her successful sale. Some bidders may have been hesitant to do business with "jerq."

DISCUSSION

In a perfect marketplace where there are multiple sellers, and where potential buyers have access to information about past sales as well as access to all current sellers, micro economics would indicate that a perfect price would emerge based on supply and demand. eBay's auction format should soon result in perfect, somewhat stable pricing.

In the case of the 168 matched auctions studied in this research, the average ending auction price was $15.15 with a range of $4.99 to $38.00. The average difference between the lowest and highest bids in matched sets was $10.76. Obviously sellers would rather be at the top of the selling range than at the bottom.

The greatest difference between the highest price bid for a particular issue ($26.85) and the lowest bid for that same issue ($0) was quite a lot given the general price range for this type of product. Since only four days separated the two auctions, the issues were matched for condition of product, and the second auction resulted in the higher price—indicating demand had not yet been met, price must be influenced by factors in addition to supply and demand. How the seller describes the product has been shown to be significant in the price the seller realizes.

The specific results of this research may not be generalizable to all product categories. It is likely that buyers of different products would prefer more or less product description -- depending on how complex or technical the product is, on whether condition is more or less important, and on how expensive the product is. If careful handling and shipping is critical to a satisfied buyer, or if an escrow service would be required in the case of very expensive items, the bidder might respond well to more shipping, handling and paying information. What this research does clearly point out is that it is worth the seller's time to do a careful comparison of the auction outcomes of products similar to the ones he or she is attempting to sell, identifying those approaches to discourse which are associated with the highest price outcomes. The cost of listing an auction on eBay is the same whether the seller receives the highest final price or the lowest in a particular category. In the case of vintage Roy Rogers comic books, the best seller is, on average, making 71 percent more than the poorest seller -- a margin with which any business would be delighted.

"Don't oversell. If you do, it's like knocking on a turtle shell trying to get him to stick his head out."

–David Crawley

REFERENCES

Berkenkotter, C. & Huckin, T. (1995). *Genre knowledge in disciplinary communication: cognition/culture/power.* Lawrence Erlbaum Associates, Publishers.

Collier, M. (2002). *Starting an eBay business for dummies.* Wiley Publishing, Inc.

Collier, M., Woerner, R., & Becker, S. (2002). *eBay for dummies, 3rd Edition.* Hungry Minds, Inc.

Powers, P. (2001). *The methodology of discourse analysis.* Jones and Bartlett Publishers.

Sinclair, J. (2001). *eBay the smart way: selling, buying, and profiting on the web's #1 auction site.* AMACOM American Management Association.

Stillar, G. (1998). *Analyzing everyday texts: discourse, rhetoric, and social perspectives.* Sage Publications, Inc.

Stoddard, S. (1991). *Text and texture: patterns of cohesion.* Ablex Publishing Corporation.

THE IMPACT OF DOMAIN-SPECIFIC STOP-WORD LISTS ON ECOMMERCE WEBSITE SEARCH PERFORMANCE

Barbara Jo White, Western Carolina University
Jenny Fortier, Multimedia Interactif
Danial Clapper, Western Carolina University
Pierre Grabolosa, Western Carolina University

ABSTRACT

Search time is an important determinant of e-service quality. Consumers may abandon slow websites prior to purchase. Much like an index in the back of a book facilitates searching, a similar index, called an inverted index, facilitates searching in a digital environment. Smaller indexes result in faster searches. One way to accelerate search time is to reduce the index size by removing common words like "the," "and," or "with." These words, called "stop words," offer little searchable meaning. Words on standard stop-word lists are frequently removed from indexes. However, e-commerce sites also contain high-frequency, low-value words that are domain-specific and should be added to standard stop-word lists. We created a corpus of over 36,000 eBay products in the furniture category and used linguistic analysis to assist in the creation of a domain-dependent stop-word list. We tested our domain-dependent stop-word list against a standard stop-word list and a control group. We experimentally generated furniture category queries and executed a random set of these queries against the three indexes and recorded the search times. A repeated-measures ANOVA was used to analyze results. As expected, the domain-dependent stop-word list resulted in the greatest reduction in search time. Implications for e-commerce sites are discussed.

INTRODUCTION

E-service quality is recognized as an important determinant of a customer's electronic commerce (e-commerce) experience. One factor contributing to a negative experience is long waiting times. In fact, consumers have a tendency to abandon web sites if waiting time becomes intolerable. Whether searching or shopping, studies have shown that consumers will wait a little longer if the system presents some sort of progress bar or hourglass to indicate that "something is happening;" however, in the absence of such features, consumers often only wait about two seconds for pages to download before leaving a website (Nah, 2004) often without making a purchase. The consumer experience often begins with a search for products or information. Shoppers on eBay log more than 400 million searches per day for any one of more than 25 million available products. Increasing the result presentation speed following a search request is one way to accelerate the search process. Generally, an inverted index is used for information retrieval purposes. Much like an index in the back of a book or facilitates searching with a list of words followed by page numbers, an inverted index facilitates searching with a list of words followed by identifiers for the electronic files. Often, some words are left out of indexes. Words like *the*, *and*, or *with* and hundreds of others rarely appear in indexes because they are used frequently in everyday language and have very little meaning. These common words, called stop-words, are removed from the index, which decreases its size. Smaller indexes are hypothesized to decrease search time. While standard stop-word lists are useful across domains, there may be other equally high-frequency low-value words within each domain.

An important e-commerce domain is that of online auctions. We chose to work with eBay auctions because of its tremendous e-commerce value—over $ 44 billion in gross merchandise sales in 2005 (www.ebay.com). Specifically, we created a corpus of the nearly 40,000 items in the furniture category of the Home& Garden section. The furniture category was chosen for several reasons: first, more than 80% of furniture buyers search online prior to purchase (Anderson, 2005); second, about 8000 pieces of furniture are sold each week on eBay, which facilitates the sale of a sofa every 17 minutes (Allegrezza, November 28, 2006); and third, the furniture industry is an important one to the state of North Carolina, where this study takes place.

This study develops a technique to create an efficient domain-dependent index that speeds up searches and empirically tests it by comparing to techniques in current practice. Typical consumer queries in the furniture domain were run against three different indexes serving as treatment groups: an index with domain-dependent stop-words removed; a second index with a smaller set of standard stop-words removed; and a control group index with no words removed. A repeated measure ANOVA was used to analyze the results.

E-SERVICE QUALITY

Search speed has been identified as a determinant of website quality, which is a major contributor to e-service quality (Cao, Zhang, & Seydel, 2005; Santos, 2003). Studies on service and website quality have utilized various methods, from focus groups (Santos, 2003) to individual ratings of websites (Lin & Lu, 2000). Each study implicates search speed as a factor contributing to website quality.

For example, website factors affecting service quality were studied using a consumer focus group, a method consistent with other marketing studies (Santos, 2003). Factors associated with e-service quality can be categorized those associated with the incubative dimension, defined as "the proper design of a Web site, how technology is used to provide consumers with easy access, understanding and attractions of a Web site (Santos, 2003, p. 238)," and those associated with the active dimension, defined as "the good support, fast speed, and attentive maintenance that a Web site can provide its customers (Santos, 2003, p. 241)." Determinants of the active (post-launch) dimension of service quality include reliability, efficiency, which refers to "the speed of downloading, search and navigation (Santos, 2003, p. 241)," support, communications, security, and incentives. Efficiency was the ranked as the one of the top two most important determinants of the active dimension (Santos, 2003). On the other hand, determinants of the incubative (pre-launch) dimension include ease of use, appearance, linkage, structure and layout, and content. Ease of use, the most important determinant of e-service quality with respect to website design (Santos, 2003), refers to "how easy the Website is for customers to conduct an external search in cyberspace and internal navigation and search within the Website (Santos, 2003, p. 239)." Clearly, search speed plays an important role in both the pre-launch design phase as well as the post-launch phase of the website.

Another study focused on consumer perceptions of websites. Consumer perceptions of search speed were captured using various survey items which loaded on a construct called *responsiveness* (Cao, Zhang, & Seydel, 2005); specific items covered website response time, whether search time or loading time was considered reasonable, and whether the website was responsive to inquiries.

Studies suggest that e-commerce sites should be designed to be fast (Lin & Lu, 2000) by reducing both search time and page loading time (Cao, Zhang, & Seydel, 2005). This study makes an important contribution to the study of the ecommerce area of e-service quality by providing a technique that can be used to increase search speed.

SEARCHING IN AN E-COMMERCE ONLINE AUCTION ENVIRONMENT

Online auctions have been an important component of consumer-to-consumer e-commerce. Major competitors include eBay and Yahoo. The value of gross merchandise sales for eBay has been growing at an average of nearly 44% per year from 2001 to 2005. The number of registered users for eBay has nearly doubled every two years. Statistics for the site appear below in table 1.

Table 1: eBay Statistics (in millions)

	2001	2002	2003	2004	2005
Net Revenue	748,821	1,214,100	2,165,096	3,271,309	4,552,401
Net Income	90,448	249,891	447,184	778,223	1,082,04
Registered Users	42,400	61,700	94,900	135,500	181,000
Gross Merchandise Sales	9,319,000	14,868,000	23,779,000	34,168,000	44,300,000

*data from annual reports from www.ebay.com

Many e-commerce search engines use inverted indexes in order to retrieve items. An inverted index operates much like the index at the back of a book. The index contains a list of distinct words contained in all documents followed by identifiers for each of the items. When consumers enter a word or phrase in a keyword search box, an index is often consulted. The search word is looked up and products associated with it are returned to the consumer. One way to make searching faster is to exclude the high frequency-words from the index. These words are called *stop-words* and they typically do not add value because they appear in most of the items to be searched. Notice in table 2, that the words *a* and *for* appear in all the documents and represent words that most consumers would not enter in a search box. These high-frequency words, if left in the index, would increase its size and increase the amount of time it takes to search through it. Indexes that are too large hamper the search process and slow it down. The key is to have an index with the right number of the right stop-words removed.

Often, search engines and information retrieval incorporate standard stop-word list of 300-600 words (Pirkola, 1998). The standard stop-word lists have often been created by aggregating a list of documents into a corpus and then producing a list of words ranked by frequency. One popular stop-word list was culled from the Brown corpus of just over 1 million total words (with just over 50,000 distinct words) that come from a diverse set of documents from domains from fiction to news to religion and humor (Kucera & Francis, 1967). By using so many domains in the creation of the stop-word list, it is believed that it would prove useful when searching a wide variety of products, documents and databases. Linguistically, the lists are composed primarily of closed class words and some open class words. While closed class words belong to categories that do not accept new members (such as determiners, articles, pronouns, possessives, particles, interjections, and possessives), open class word categories (such as nouns, verbs, adverbs and adjectives) do admit new members (Klammer & Shulz, 1982). The standard stop-words are removed from the index. A sample inverted index, with no stop-words removed, for four products from appears below in table 2.

Table 2: Sample inverted index entries

Four Product Descriptions	Word	Item #1	Item #2	Item #3	Item #4
#1 5ft. Fuf Sac Bean Bag Love Seat Foof	5	#1	#2	#3	
Chair Beanbag	a	#1	#2	#3	
	auction	#1	#2		
This listing is for a new Foof Wedge Lounger. The long-lasting material can be re-foofed. Thanks for bidding. Please see my other auction items.	bag	#1	#2	#3	#4
	barbie			#3	
	be	#1	#2	#3	#4
	bean	#1	#2	#3	#4
	beanbag	#1	#2		#4
	bidding	#1		#3	
	can	#1			

Table 2: Sample inverted index entries

Four Product Descriptions	Word	Item #1	Item #2	Item #3	Item #4
#2 SNUGGLE POOF/FOOF/BEAN BAG CHAIR	chair	#1	#2	#3	#4
	choice		#2		#4
Up for auction is a chair you'll snuggle up in.New--comes in 5 colors—your choice. Thanks for looking. Free shipping.	colors		#2		#4
	comes		#2		
	foof	#1	#2		
	for	#1	#2	#3	#4
	free		#2		
	ft.	#1		#3	
#3 My Barbie Bean Bag Chair-includes shipping.	fuf	#1			
	in		#2		
Looking for a new bean bag chair with washable material? Then, you'll want to be bidding on this 5 ft. Barbie bean bag chair.	includes		#2		
	is	#1	#2		
	items	#1			
	listing	#1			
	long-lasting	#1			
	looking		#2	#3	
#4 A Bean Bag / Beanbag Chair	lounger	#1			
	love	#1			#4

Table 2: Sample inverted index entries

Four Product Descriptions	Word	Item #1	Item #2	Item #3	Item #4
You'll love this new bean bag chair. Your choice for colors. (No shipping)	material	#1		#3	
	my	#1		#3	
	new	#1	#2	#3	#4
	no				#4
	on			#3	
	other	#1			
	Please	#1			
	proof		#2		
	re-foofed	#1			
	sac	#1			
	seat	#1			
	see	#1			
	shipping	#1	#2	#3	#4
	snuggle		#2		
	thanks	#1		#2	
	the	#1			
	then			#3	
	this	#1		#3	#4
	to			#3	
	up		#2		
	want			#3	
	washable			#3	
	wedge	#1			

Table 2: Sample inverted index entries					
Four Product Descriptions	Word	Item #1	Item #2	Item #3	Item #4
	with			#3	
	you'll		#2	#3	#4
	your		#2		

We used the Penn Treebank part-of-speech (POS) tagger (Santorini, 1990) to tag the words in several standard stop-word lists. The Penn Treebank POS tagger, which is freely available, tags words with one of 36 tags (Santorini, 1990). The 36 Penn Treebank categories were collapsed into a smaller set of nine categories used in this study. The nine collapsed categories, examples, and Penn Treebank categories are shown in table 3 below.

Table 3: Parts of Speech Categories, Examples and Penn Treebank Counterparts		
POS Category	Examples	Penn Treebank Categories
Adjectives	better, best	Adjectives, including comparative and superlatives adjectives
Adverbs	quickly, well, where	Adverbs, including comparative and superlative, and wh-adverbs
Nouns	box, shipment, she	Nouns, including existential there, common, proper, and pronouns
Verbs	bid, pay,	Verbs, including modals, gerunds, and present particles
Cardinal Numbers	four, 1960	Cardinal Numbers

Table 3: Parts of Speech Categories, Examples and Penn Treebank Counterparts		
POS Category	Examples	Penn Treebank Categories
Conjunctions	and, but, nor	Conjunctions, including subordinating and coordinating
Determiners	the, which, that	Determiners, including articles, predeterminers and wh-determiners
Exclamations	uh, L@@K!	Exclamations and interjections
Other	$, £	Symbols, to, list item markers, and foreign words

The three stop-word lists that were tagged included the following: the stop-word list from the SMART information retrieval system produced by IBM; the stop-word list from the British National Corpus; and the Francis and Kucera stop-word list, created from the Brown corpus mentioned above. Tagging three standard stop-word lists shows that there are indeed differences between these "standard' lists. For example, while all three lists have approximately the same amount of conjunctions, cardinal numbers and determiners, the BNC and Kucera lists over twice as many nouns as the SMART list. In addition, the SMART list has nearly twice as many adjectives as the other two lists. Statistics for the three lists appears in table 4 below.

Table 4: Standard Stop-word list Statistics			
Column1Heading	SMART	BNC	Francis and Kucera
Number of Words	600	426	425
Adjectives	8.75%	13.15%	11.53%
Adverbs	26.26%	13.38%	14.59%
Nouns	14.26%	32.86%	38.35%
Verbs	26.42%	20.66%	17.88%

Table 4: Standard Stop-word list Statistics			
Column1Heading	SMART	BNC	Francis and Kucera
Cardinal Numbers	1.62%	1.64%	1.18%
Conjunctions	10.21%	9.39%	9.65%
Determiners	2.76%	3.29%	3.53%
Pronouns	7.62%	5.16%	2.82%
Other	2.11%	0.47%	0.47%

In place of standard stop-word lists, however, studies suggest that domain-dependent stop-word lists might be more efficient (Harman, 1994). Domain-dependent stop-word lists could be created through the modification of one of the standard stop-word lists (Natt och Dag, Regnell, Carlshamre, Andersson, & Karlsson, 2002), which could be accomplished by producing a list of high frequency words and then removing those words considered valuable while leaving the words that are considered of low value (Harman, 1994). For example, notice in table 2 again that there are also high-frequency words such as *new*, *shipping* and *bean bag chair* which also appear in all documents. These words are high-frequency words, yet they are more specific to the eBay domain. The word *shipping* should be included in a stop-word list but *new* and *bean bag* chair should not (Surendran, Platt, & Renshaw, 2005; White, Lindblom, & Conlon, 2003).

Domain-dependent stop-word lists have been created manually by modifying an initial list created from a variety of methods, from methods including frequency counts (Banerjee & Rudnicky, 2006; Chang, Raychaudhuri, & Altman, 2001; Kamvar, Oliver, Manning, & Altman, 2003), to methods including frequency counts and word length (Donaldson et al., 2003), to methods that researchers do not describe (Fan & Kambhampati, 2005; Forman, Kirshenbaum, & Suermondt, 2006). Another novel method of creating a domain dependent stop-word list has included some form on linguistic analysis (Surendran, Platt, & Renshaw, 2005; White, Lindblom, & Conlon, 2003). In research designed to facilitate searching and categorization, frequency counts were used to determine words left out of the index. In addition, a part of speech tagger was used to leave noun phrases in the index (Surendran, Platt, & Renshaw, 2005). The effectiveness of these lists has been evaluated by personal inspection (Surendran, Platt, & Renshaw, 2005) or by reporting index size and, anecdotally, response time statistics (White, Lindblom, &

Conlon, 2003). This is the only study to utilize a large corpus of e-commerce products and then to experimentally test the efficiency of using a repeated measures design.

METHOD

The method we used to create our domain dependent stop-word list included the following steps: creating a corpus of items from the eBay auction website; cleaning the corpus of unwanted punctuation and symbols; using linguistic analysis to determine parts of speech in different areas of the corpus titles, descriptions, or titles and descriptions; determining the potential stop-word list by using token frequency counts; and altering the potential list based on the linguistic analysis of the product titles and descriptions in the corpus of products.

CREATION AND CLEANING OF THE EBAY CORPUS

Only recently has eBay allowed developers to retrieve data from their store of products. As of June, 2005, there were nearly 15,000 developers registered with eBay's developer network (http://developer.ebay.com/) using Application Programmer Interface (API) calls to interact with eBay data. Developers are allowed 10,000 calls to the database per month (www.ebay.com). First, products from the furniture category were collected and loaded into a database. Then, after the corpus was created, it was cleaned of unwanted characters. The unwanted characters included javascript, html, and various symbols and punctuation marks. A program was written to remove unwanted characters from the eBay corpus of products. From the cleaned text, three text files were created: product titles; product descriptions; and product title and descriptions. In addition, separate text files were created for each of the 36,075 product titles as well as for each of the separate product titles with its associated descriptions.

ALGORITHM TO CREATE THE DOMAIN-DEPENDENT STOP-WORD LIST

Text files of the eBay corpus product data were analyzed and used to create the domain-dependent stop word list. The first step was to develop a potential list of stop-words. This was accomplished by using the CMU-Cambridge Statistical Language Modeling Toolkit (Clarkson & Rosenfeld, 1997) to perform frequency counts for each distinct word in the corpus. The top-ranked word, *the*, had a frequency count of 665,761, which amounts to an average of nearly 18.5 occurrences, or tokens, per eBay listing. Data for the corpus appears in table 5 below.

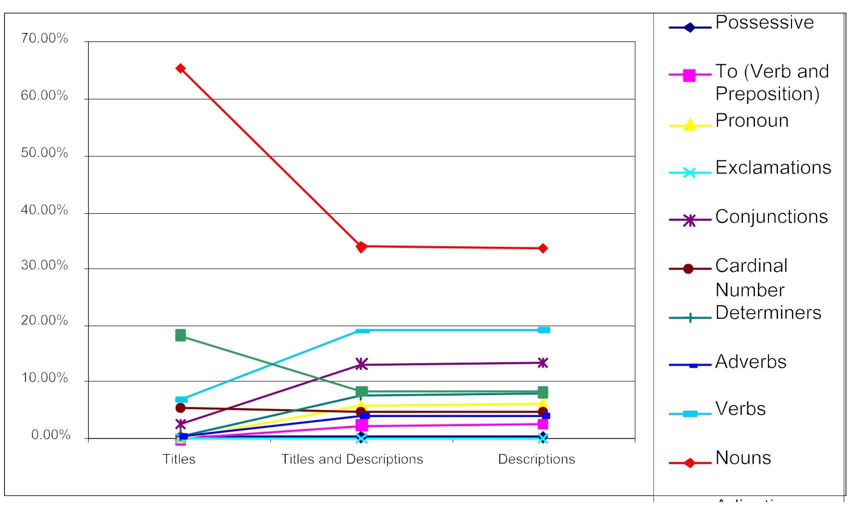
Table 5: Statistics of the eBay Corpus	
	Data
Number of Items	36,075 items
Total Word Count (Tokens)	19,696,789 tokens
Avg. Words Per Title	8 words
Avg. Words Per Description	538 words
Total Distinct Words	65,139 words

Words with a frequency count of 75 serve as the starting point for the domain dependent stop-word list. The result was a list of 7,027 words, which represented approximately 10 % of the distinct words and 94% of the total tokens, or word occurrences.

To determine which words to remove from the potential stop-word list, the first step involved conducting a linguistic analysis of eBay product items using a part of speech (POS) tagger (White, Clapper, Noel, Fortier and Grabolosa, 2007). Trained on particular data sets, taggers have been shown to perform differently when used to tag documents in a different domain; however, accuracy scores across two taggers and four training data sets were within less than 1% of each other (Boggess et al., 1999). We used the Penn Treebank POS Tagger to tag words in the eBay corpus with one of 36 tags (Santorini, 1990). Specifically, we tagged three parts of each product listing: the titles, descriptions, and titles and descriptions taken together. These components were selected because it was believed that titles and descriptions would show different patterns with respect to parts of speech. In addition, using keyword searches on the title field is the default method of searching on eBay. Searching more than just the title field requires the consumer to activate a check box indicating the desire to search titles and descriptions. Data for selected part-of-speech tags by sub-sections of the corpus appear in table 6 below.

Table 6: eBay Statistics			
Column1Heading	Titles	Titles and Descriptions	Descriptions
Adjectives	18.20 %	8.36%	8.22%
Adverbs	0.48 %	4.01 %	4.06 %
Nouns	65.53 %	33.94 %	33.48 %
Verbs	6.83 %	19.12%	19.29%
Cardinal Numbers	5.22 %	4.77%	4.77 %
Conjunctions	2.77 %	13.32%	13.47 %
Determiners	0.46 %	7.77 %	7.87 %
Exclamations	0.00 %	0.07 %	0.07 %
Other	0.43 %	2.72 %	2.76 %

The graph below illustrates the above data—the part of speech tagging for the 20 million words of the 36,075 eBay product listings. It can be removed from the paper if tables create problems for printing purposes



It is clear that the major components of the product listings do show different patterns with respect to part-of-speech tags. For example, nouns and cardinal numbers are more plentiful in titles compared to descriptions whereas verbs and adverbs are more common in descriptions compared to titles (White, et al., 2007).

The algorithm for the domain-dependent stop-word list started with the list of words occurring more than 75 times. Then, nouns, cardinal numbers and adjectives are removed from the list. The resulting domain-dependent list of stop-words consisted of 1457 words that included verbs (74%), adverbs (17%), conjunctions and prepositions (6%), and determiners (1%). The algorithm for the domain-dependent stop-word list appears below in table 7.

Table 7: Algorithm for the eBay Domain Dependent Stop-word list -
For each word in the corpus with a frequency greater than 75   Remove nouns, adjectives and cardinal numbers End for

EVALUATION OF THREE INDEXES USING SEARCH QUERIES

To evaluate the effectiveness of the domain-dependent stop-word list, we used a repeated measures design. We created three indexes for the eBay corpus of nearly 40,000 products: one one with no stop words removed; one with the domain-dependent list of stop words removed; and one with a standard set of stop-words removed. A set of random queries was submitted to each index and response times recorded. Data were analyzed with a repeated measures ANOVA. It is hypothesized that a smaller index yields faster response times. Details of the evaluation process follow.

The eBay corpus was indexed using a modified perl script (Slesinsky, 1997).The Unix-based software allows for the creation of inverted indexes. In addition, for each index created, it is possible to enter stop-word lists in the form of a text file of words that should be removed from the index. The Characteristics of the three indexes are shown below in table 8.

Table 8: Characteristics of the Three Indexes			
	EBay Index	Standard Stopword Index	Control (No Stopword) Index
Inverted Index FileSize	51.7 MB	56.4 MB	69.7 MB
Building Time	14,463 seconds	18,655 seconds	38,937 seconds

After creating the three indexes, they were evaluated with a set of random queries. To generate the random queries, students served as subjects and generated queries in response to a set of image prompts. The use of students to generate random queries is a method that has been used successfully in prior studies (White, Posey, & Conlon, 2004). Characteristics of the subjects appear in table 9.

Table 9: Characteristics of Query Generation Subjects		
	Male	Female
Number	16	9
%	64	25

Images were collected from www.ebay.com. About 60 pictures were collected randomly from the “Furniture” subcategory under the superordinate “Home & Garden” category. Images were excluded from the set if the background was busy, or if multiple items were present in the image. Multiple items in images typically occur when the “item” is a set, as in a set of bedroom, living room or dining room furniture; however, multiple items can also occur in images which prominently feature accessory items, such when a bookcase features a television, lamp or other accessory. Multiple items or busy backgrounds make it difficult for subjects to identify the object serving as the prompt for the query phrase. Because the corpus contains all the items in the furniture category, it was decided to include image prompts that represented similar percentages to the subcategories of the furniture category shown in table 10. The queries used in this study were selected while maintaining these percentages, essentially creating a stratified random sample.

Table 10 Percentage of Items in Subordinate categories of the Furniture Category					
Category	2/11/2007	2/18/2007	2/25/2007	Average	Images in Study
Living Room Furniture	47.98 %	47.63 %	47.73 %	47.78 %	46.67%
Bedroom Furniture	21.15%	21.18 %	21.16 %	21.16 %	30.00%
Dining Room Furniture	12.19 %	12.09 %	12.11 %	12.13 %	10.00%
Children’s Furniture	7.48 %	7.48 %	8.01 %	7.65 %	3.33%
Office Furniture	7.22 %	7.28 %	7.28 %	7.26 %	6.67%
Kitchen Furniture	6.47%	6.70 %	7.23 %	6.8 %	3.33%

Subjects recorded their data on a spreadsheet in which they first recorded basic demographic data common to language variation studies, such as gender and where they grew up and if their parents were from that area. Then, subjects viewed each image in color, and typed their initial search term or phrase, so as to reduce the effects of illegible writing. Query terms were aggregated for each image. Query data appears below in table 11.

Table 11			
Gender	Male	Female	Average
Number of Words Per Query	2.54	3.07	2.81
Distinct Terms Per Picture	13	8	10

From the population of 1500 queries, and while maintaining the percentages shown in table 10, thirty queries were randomly selected to test the three indexes. To reduce the effects of computer system variation, each query was first run 20 times against each of the three indexes. The time in milliseconds from the start to the finish of the process was recorded for each run and then, an average was computed. Query times varied from less than 65 milliseconds to over 7200 milliseconds. In effect, the average search time for the query itself becomes the subject in our repeated measures design and each index serves as a treatment. The average time for each index appear below in table 12.

Table 12 Search time means for the three indexes for the three categories of queries

Column1Heading	eBay Index	Standard Stopword Index	Control Index	Avg. Search Time
Query Times (in milliseconds), N=33	903.64	954.91	1081.52	980.02

The queries tested using the domain-dependent index performed nearly 5.6% faster than the same queries run against the index with a standard list of stop-words removed, and nearly 20% faster than the control index with no stop-word list. Search times were analyzed using a repeated measures analysis of variance (ANOVA). The index (eBay vs. Standard Stopword vs. Control) served as a within subjects-factor. Mauchley's test of sphericity was significant so the assumption of sphericity, which is a homogeneity of variance type of assumption that deals with equality of variances of differences between treatments, was not met. To deal with this violation the Huynh-Feldt correction was applied in order to adjust the degrees of freedom. The effect of the index size was significant [$F(1.061, 33.945)=3.761$, $p=.059$] at an alpha level of .10. Paired T-Tests were also conducted. These simple contrasts showed that the times (in milliseconds) for queries run against the domain-dependent index were significantly faster than both those run against the index with the standard set of stop-words removed [$t(32)=1.996$, $p=.055$] and those run against the index with no stop-words removed [$t(32)=1.999$, $p=.054$].

CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH

This study makes several important contributions to the ecommerce field. First, the study uses a large corpus of nearly 20 million words comprising just over 36,000 products from eBay, an important ecommerce website. Second, it considers the language that consumers use in the online auction marketplace when implementing techniques that accelerate website searches as the technique is based on a linguistic analysis of consumer-provided product listings.

There are several limitations with this study as well. First, though the algorithm was created using data from tagging the entire corpus of nearly 20 million words, the algorithm was implemented using the tagged set of nearly 70,000 distinct words. Words tagged in context often carry different parts of speech from words in a list. For example, a noun that appears before another noun in a noun phrase will be tagged as an adjective in the context of the whole sentence, but as a noun if tagged as a single word. Even though a few words on the stop-word list may have been affected, it is expected that this limitation would not affect the basic decisions to remove adjectives, nouns and cardinal numbers from the stop-word list or the outcome of the study. Second, the 1800 queries were run by hand and data collected by hand. There exists the possibility of error in the absence of a well-tested automated program to run the queries and collect the data. Third, words that are misspelled are may be tagged incorrectly. For example, the eBay corpus has the word *chandelier* spelled three ways tagged twice as a noun and once as an adjective.

The corpus provides a rich data set for future research. One area of study could include text-based language variation in the area of online furniture shopping. Another area of research inovlves optimization techniques that could be used to determine the optimum number of stop-words to remove from a domain-specific eBay index. In addition, optimization techniques can then be used to determine the smallest number of products that would produce the optimum list of stop-words. The resulting index could be tested against a very large set of consumer-generated queries.

REFERENCES

Allegrezza, R. (November 28, 2006). Stop crying and play hardball. *Furniture Today* Retrieved February 12, 2007, from http://www.furnituretoday.com/blog/380000038-November-2006.html

Anderson, K. (2005). Web impacts furniture buying habits. *Furniture Today* Retrieved February 12, 2007, from http://www.furnituretoday.com/article/CA6375724.html

Banerjee, S., & Rudnicky, A. I. (2006). *You are what you say: Using meeting participants' speech to detect their roles and expertise.* Paper presented at the Analyzing Conversations in Text and Speech (ACTS) Workshop at HLT-NAACL 2006, New York City, New York.

Boggess, L., Hamaker, J. S., Duncan, R., Kimek, L., Wu, Y., & Zeng, Y. (1999). *A comparison of part of speech taggers in the task of changing to a new domain.*

Paper presented at the 1999 International Conference On Information Intelligence and Systems, Bethesda, MD, USA.

Cao, M., Zhang, Q., & Seydel, J. (2005). B2C e-commerce web site quality: An empirical examination. *Industrial Management & Data Systems, 105*(5), 645-661.

Chang, J. T., Raychaudhuri, S., & Altman, R. B. (2001). *Including biological literature improves homology search.* Paper presented at the Proc. Pacific Symposium on Biocomputing (PSB).

Clarkson, P., & Rosenfeld, R. (1997). CMU-Cambridge Statistical Language Modeling Toolkit. (version 2.05). Retrieved Month ##, 2006, from http://mi.eng.cam.ac.uk/~prc14/toolkit.html

Donaldson, I., Martin, J., de Bruijn, B., Wolting, C., Lay, V., Tuekan, B., et al. (2003). PreBind and Textomy- mining the biomedical literature for protein-protein interactions using a support vector machine. *BMC Bioinformatics, 4*(11).

Fan, J., & Kambhampati, S. (2005). A snapshot of public web services. *SIGMOD Record, 34*(1), 24-32.

Forman, G., Kirshenbaum, E., & Suermondt, J. (2006). *Pragmatic text mining: Minimizing human effort to quantify many issues in call logs.* Paper presented at the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia, PA, USA.

Harman, D. (1994). NIST Interagency Report 4873: Automatic indexing. Retrieved February 2, 2007, from http://www.-nlpir.nist.gov/works/pubs/ir4873.html

Kamvar, S. D., Oliver, D. E., Manning, C. D., & Altman, R. B. (2003). Inducing novel gene-drug interactions from the biomedical literature. *Bioinformatics, 1*(1), 1-8.

Klammer, T. P., & Shulz, M. (1982). *Analyzing English grammar*. Needham Heights, Massachusetts: Allyn and Bacon.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day Amercian English* Providence, Rhode Island: Brown University Press.

Lin, J. C.-C., & Lu, H. (2000). Towards an understanding of the behavioral intention to use a web site. *International Journal of Information Management, 20*(3), 197-208.

Nah, F. F.-H. (2004). A study on tolerable waiting time: How long are Web users willing to wait? *Behavior and Information Technology, 23*(3), 153-163(111).

Natt och Dag, J., Regnell, B., Carlshamre, P., Andersson, M., & Karlsson, J. (2002). A feasibility study of automated natural language requirements analysis in market-driven development. *Requirements Engineering, 7*(1), 20-33.

Pirkola, A. (1998). *The effects of query structure and dictionary setups in dictionary-based cross-language information retrieval.* Paper presented at the 21st annual international ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Australia.

Santorini, B. (1990). *Part-of-speech tagging guidelines for the Penn Treebank Project*: Technical report MS-CIS--47, Department of Computer and Information Science, University of Pennsylvania.

Santos, J. (2003). E-service quality: A model of virtual service quality dimensions. *Managing Service Quality, 13*(3), 233-246.

Slesinsky, B. (1997). Roll your own search engine. Retrieved February, 10, 2007, from http://www.webmonkey.com/webmonkey/97/16/index2a.html

Surendran, A. C., Platt, J. C., & Renshaw, E. (2005, July). *Automatic discovery of personal topics to organize email.* Paper presented at the 2nd Conference on Email and Anti-Spam, Stanford University, USA.

White, B. J., Lindblom, T., & Conlon, S. (2003, November 22-25). *Decreasing query response times using statistical and syntactic analysis to reduce the size of the inverted index.* Paper presented at the 34th Annual Meeting of the Decision Sciences Institute (SEDSI), Washington, D.C.

White, B. J., Posey, J. H., & Conlon, S. (2004, August 5-8). *When search engines "speak" your language: The role of communication accommodation theory in personalization systems.* Paper presented at the Proceedings of the Tenth Americas Conference on Information Systems, New York, New York.

White, B. J., Clapper, D., Noel, R., Fortier, J.,& Grabolosa, P. (2007). Understanding persuasive online sales messages from eBay auctions. *Business Communication Quarterly, 70*(4), 482-487.

Allied Academies

invites you to check our website at

www.alliedacademies.org

for information concerning

conferences and submission instructions