

COMP6049 Quantitative and Qualitative Methods

Week 1

Methods and Methodology

Handout and in class exercise:

Exercise: research abstracts

Skim read the abstract you have been given and briefly discuss with the person next to you the following

- What kind of science is this ?
- What methods were used ?
- Did you like this paper and want to read more?



You can look at the full set of examples on edshare accessed via the module pages 10

The following examples of journal articles are designed to illustrate some part of the wide range of different approaches which are adopted across different academic areas.

Participants are invited to skim read one of the articles, and then to discuss any observations on the content and the style.

During the class we will discuss disciplinary differences with reference to personal experiences and the variability of approaches across the examples.

DAWN OF THE DEATH OF DISTRIBUTED DENIAL OF SERVICE: HOW TO KILL ZOMBIES

LILIAN EDWARDS*

TABLE OF CONTENTS

I. INTRODUCTION TO THE PROBLEM	23
A. <i>Prevalence and Effects of DDOS</i>	28
B. <i>Who Suffers DoS/DDOS and Why</i>	32
II. LEGAL RESPONSES TO DoS AND DDOS.....	35
A. <i>Criminal Law</i>	36
B. <i>DoS and Intent: the "Possessed by Aliens" Defense</i>	41
C. <i>Civil Law</i>	43
1. <i>Sue the Zombies?</i>	46
2. <i>Sue the Software Writers?</i>	51
III. SECURITY IS FOR EVERYONE, NOT JUST FOR CHRISTMAS .	56
A. <i>Targets</i>	58
B. <i>ISPs</i>	59

I. INTRODUCTION TO THE PROBLEM

According to Wikipedia, the on-line encyclopedia, a "*denial of service*" or "DoS" attack is:

an attack on a computer system or network that causes a loss of service to users, typically the loss of network connectivity and services by consuming the bandwidth of the victim network or overloading the computational resources of the victim system.¹

In layman's terms, a DoS attack has the effect of making a computer system—whether it is operated by a business, a public sector resource, or an individual—unable to supply its services to clients. CERT, based at Carnegie-Mellon University and the leading center for Internet security, categorizes a "denial-of-service" attack as "an explicit attempt by attackers to prevent

* LLB (Hons)(Glas), LLM (Cantab), MSc in IT(York); Co-Director, Arts and Humanities Research Council Centre for Research into Intellectual Property and Technology Law, University of Edinburgh, and Chair of Law, University of Southampton. The title of this essay derives from perhaps the most famous zombie film of recent years, George Romero's *Dawn of the Dead* (1978), the source of numerous parodies, including Edgar Wright's excellent *Shaun of the Dead* (2004). As will be discussed below, "zombie networks" are the proximate cause of the blight of distributed denial of service (DDOS), and are increasingly the main conduit for distribution of spam, malware, and phishing attacks.

¹ Wikipedia, *Denial-of-service attack*, <http://en.wikipedia.org/wiki/DDOS> (last visited Feb. 26, 2006).

E-PROCUREMENT AS A DEVELOPMENT IMPERATIVE FOR SMALL ISLAND STATES IN THE SOUTH PACIFIC

MOHAMMED AHMADU

- [INTRODUCTION](#)
- [I LEGAL MECHANICS OF INTERNET TECHNOLOGY](#)
- [II LOCAL PROCUREMENT SCENARIO](#)
- [III SYNOPSIS OF LOCAL ICT FRAMEWORKS](#)
- [IV REGIONAL ICT PERSPECTIVE](#)
- [CONCLUSION](#)

E-PROCUREMENT AS A DEVELOPMENT IMPERATIVE FOR SMALL ISLAND STATES IN THE SOUTH PACIFIC

MOHAMMED AHMADU*

* Mohammed L Ahmadu LL.M (Warwick) ACIS (UK) Associate Professor of Law, University of the South Pacific; doctoral candidate in law.

Abstract Conventional legal principles that support contractual relationships must be adapted to keep abreast of developments in the evolving ICT revolution (digital technology and the Internet) and e-commerce. Small island states — and developing countries in general — cannot afford to lag behind: to position their economies in the global trading system they must embrace e-procurement as a development imperative. This article attempts to make the case that the economic and technological vulnerability of small island states in the South Pacific will be minimised, and their capacity to effectively participate in global trade enhanced, if they establish a functional and fully integrated legal and technological framework to support electronic-based transactions nationally and regionally.

INTRODUCTION

The increasing pace of globalisation and advancements in technology are stretching the traditional boundaries of contract and commerce, so much so that the conduct of electronic commerce by means of online transactions is fast becoming a worldwide phenomenon. While small South Pacific island states wish to be part of this trend, they are nonetheless vulnerable in a number of ways. These include fragile ecology, comparatively smaller economies with low absorptive capacities, and the oft-perceived geographical disadvantage associated with remoteness from major global and trading centres.

To overcome these challenges the use of the Internet and associated technologies, anchored and operating in a suitable legal framework, is now necessary. These will assist in converting disadvantages into strengths rather than continually looking at the small island states as helpless. With a suitable legal framework and functional ICT infrastructure, particularly in the area of e-procurement or commerce, small island states in the South Pacific will be in a position to open up their economies to much-needed foreign investments and equity or joint venture capital. At the same time these states will be able to electronically market their natural resources or products regionally or globally, without having to engage expensive middle dealers or intermediary services. There are innumerable benefits, both potential and actual, which flow from the electronic procurement of goods, services or construction.

procurement of goods, services or construction.

To set the scene, a brief discussion on the techno-legal basis of the Internet will be undertaken. Internet technology provides the primary mechanism through which widespread electronic transactions can be conducted with relative push-button ease. The technology is particularly relevant to the situation of small island states of the South Pacific because of their remote geographical location and limited infrastructure. The experiences of Malaysia and Singapore will also be discussed in order to show the relevance of their ICT practices, in the field of online transactions, to small South Pacific island states. In view of their developing-country status and their location in the Asia Pacific region, Malaysia and Singapore could serve as ideal legal and technical models to small island states.^[1]

This article simply aims at examining the legal possibility of incorporating an online transaction or contract as a platform for the procurement of goods, services or construction in small Island states of the South Pacific.

I LEGAL MECHANICS OF INTERNET TECHNOLOGY

Electronic commerce and the use of Internet-based arrangements to procure goods or services^[2] illustrate a marriage of convenience between law and technology. While the medium or vehicle through which such transactions are conducted is provided by technology, the rules for validating and enforcing the transactions are supplied by the law. In this area, it is impossible for the law to exist independent of the technology, and vice versa.

In the conventional legal sense, however, the validity and enforceability of a transaction made or contract entered into by parties depends on the existence of offer, acceptance, consideration and intention to create legal relations. The law assumes that the parties will be able to establish the existence of a contract whether orally, in writing, by conduct or partly through any of these three means. Furthermore, the rules in relation to these four cardinal elements of contract are fairly settled law. As the use of the Internet to transact through the formation of online contracts raises fundamental legal issues, a brief exposition of how the Internet works is provided here.

The Internet provides an open channel of electronic communication covering vast distances, using complex technology and equipment to transmit or exchange data, sound or images between multitudes of computers all connected in a global communications network. In this regard, the Internet has been appropriately described as communications technology.^[3]

Information sent from one computer to another via the Internet is transported through several servers at random by using the most efficient route until it finally reaches its intended destination. The information is broken into 'packets' for transit at the point of despatch and reassembled into its original form at the point of receipt. For the information to be transported through the network, a protocol must be used, the most important being the Transport Control Protocol (TCP).^[4] The Internet Protocol (IP) determines what server should receive the 'packet of information' for onward transmission. To access the Internet via any website^[5], the user must indicate the correct Uniform Resource Locator (URL) to visit.

The use of this technological infrastructure means that at any given time, multiple copies of packets of information are temporarily copied by different servers in different locations worldwide. The physical location of each server depends on the routing. Emails are regulated by means of other application protocols, namely the Simple Mail Transfer Protocol (SMTP) and the Post Office Protocol (POP).^[6]

Intercultural Communication on Web sites: A Cross-Cultural Analysis of Web sites from High-Context Cultures and Low-Context Cultures

Elizabeth Würtz

Department of Design, Communication and Media
IT University Copenhagen

The aim of this study is to explore and identify the strategies used by high-context cultures in utilizing the Internet—a largely low-context medium—for communication and marketing purposes. It is hypothesized that individuals in high-context cultures are more likely to adopt the visual effects offered by the Internet to convey their messages efficiently than their low-context counterparts. How might high-context cultures make the most of the potentials offered by the Internet generation of today? Assuming that visual communication is a high priority in the design of high-context Web sites, how do the visual methods used on Web sites vary according to the communication styles in different cultures? Using Hall's high- and low-context dimensions as the main parameters, an exploratory analysis of McDonald's Web sites identified five different strategies by which visual communication is used to support high-context communication traits.

doi:10.1111/j.1083-6101.2006.00013.x

Introduction

A popular cultural framework was proposed by Edward Hall (1976, 2000), in which he stated that all cultures can be situated in relation to one another through the styles in which they communicate. In some cultures, such as those of Scandinavians, Germans, and the Swiss, communication occurs predominantly through explicit statements in text and speech, and they are thus categorized as low-context cultures. In other cultures, such as the Japanese and Chinese, messages include other communicative cues such as body language and the use of silence. Essentially, high-context communication involves implying a message through that which is not uttered. This includes the situation, behavior, and paraverbal cues as integral parts of the communicated message.

These differences in communication styles across cultures are expected to pose challenges to the ways in which Web sites communicate their messages most optimally. How do Web sites created for a target group in a high-context culture differ from those created for low-context culture audiences? What strategies do high-context cultures use to compensate for lack of context on Web sites that were created in cultures where the norm is to communicate in a manner that is high in content and low in context?

This article presents a set of preliminary results describing the tendencies by which communication through Web sites is adapted to various cultures. The underlying premise of the article is that when customizing a Web site to appeal to a different culture it is not enough merely to translate the text; the overall communication strategy should be appropriate to the audience as well. The study identifies whether and how variables that characterize high- and low-context cultures are reflected on Web sites. It further attempts to draw parallels between face-to-face communication and communication on the web by looking at communication rules and patterns in high-context cultures and comparing and contrasting them to the communication style of Web sites in low-context cultures.

Cultural Awareness in Web Design

At about the same pace as the popularity of the Internet increased, visions flourished of the World Wide Web as a tool for bringing the world together. The marketing world in particular quickly embraced the Internet as an ideal medium for reaching beyond domestic markets in order to disseminate products to hitherto foreign markets. The localization strategies for this may have been simple at first—first generation Web sites, which were simple and text-based, needed a mere translation to be cross-cultural. However, the development of Flash and the implementation of video and sound have brought new potentials to the Web and set new standards for efficient and effective Web communication.

Nowadays a Web site is not just a collection of text; it is a conglomerate of images, multimedia, interactive features, animated graphics, and sounds. From a marketing-strategic perspective, a company that defines itself as cross-culturally aware knows (or should know) that creating appealing and efficient Web sites for other cultures is no longer just a matter of language and modification of time- and date-formats. Cross-cultural Web design nowadays requires dealing with design issues that include culture-specific color connotations, preferences in layout, animation, sounds, and other effects that are characteristic of today's generation of Web sites.

In order to do this successfully, the designer must study the target group of the Web site. While user participation is ideal in the designing process, a study of the design elements prevalent in the culture may also provide the Web designer with some useful guidelines. Values and behavior indoctrinated through cultural influences may be reflected in design practices.

By understanding how communication styles may be reflected on Web sites, we come a step further towards identifying, and subsequently realizing the potentials of,

Cite as: White, M.M., & Mann, B.L. (2009). Blurring our real and virtual worlds: Canadian and worldwide legal issues arising from MMORPGs, *forthcoming*.

**Blurring Our Real and Virtual Worlds:
Canadian and Worldwide Legal Issues Arising From MMORPGs**

Matthew M. White¹ and Bruce L. Mann²

June 14, 2009

ABSTRACT

In recent years, more and more people have become members of virtual online worlds through the promulgation of massively multiplayer online role playing games (MMORPGs). As of December 28th, 2008, World of Warcraft, a popular MMORPG, reached 11.5 million players – a figure that would make the fictional world of Azeroth more populated than Cuba.³ It is possible that by 2011, four out of every five people who use the Internet will work or play in a virtual world.⁴ With so many players investing time and money into these online games, legal issues have begun to arise that draw close parallels between game rules and real world laws. Issues such as individual rights and character rights as designated by in-game End User License Agreements (EULAs), ownership of in-game property, gold farming and child labour, and criminal prosecution and jurisdiction for in-game crimes. This paper critically examines the close proximity of in-game legal issues to legal issues faced in the real world, and argues that as more people begin to adopt these technologies, the lines between virtual and real will become increasingly more difficult to discern.

Blurring Identities and Citizenship

Insomuch as one expects representational powers, such as the right to vote, to freedom of speech, to ownership of land, and so on, as recompense for obeying his or her Canadian civic duties – such as taxation – many apply to other countries of residence, one might similarly expect to be granted equivalent rights whilst taking residence in a virtual world and paying taxation by way of in-game and real life monetary fees. Individuals are sometimes as patriotic about the lands inhabited by their character players or avatars as they are about their home country, shouting “for the Horde!” at World of Warcraft conferences, and sporting Azeroth-emblazoned regalia during everyday life; something akin to wearing the maple leaf across one’s chest. The legal issues arising from this trend are twofold; initially – what rights of citizenship do participants in the virtual world of a game hold, which are eschewed, and under what authority? Additionally, to what extent, and from what precedent can fees and subscription payments, as well as an active participation in a game in which one participates as co-author of the sum experience through role-play and character creation be justifiably sufficient in granting the usual rights of citizenship afforded to individuals reciprocally from their civic duties.

To be sure, gaming companies, as proprietors of virtual worlds, ascribe the same legal duties to the citizens of their virtual worlds as do real countries across the world. Insomuch as a citizen must

obey the laws of his/her country, lest s/he face punitive measures, so too must an avatar. Perhaps a less cited obligation is that of civic participation. While it is not uncommon for citizens of a nation to face jail time for avoiding civic participation, such as jury duty, it is seldom cited in reference to virtual worlds.

MDY Industries v Blizzard Entertainment, Inc. Blizzard, the maker of *World of Warcraft* is suing Michael Donnelly, creator of a program called *MMO Glider* that allows players to gain levels and in-game wealth through playing the game automatically while the player is away from the keyboard. Blizzard argues that this violates their EULA, which enforces participation while the character is logged in. The important feature here is that the human player must be present, not simply the avatar logged-in, in order for the EULA to be upheld.⁵ From this case, a particularly interesting precedent is emerging – if an avatar without a human player actively monitoring and controlling its actions does not hold the right, despite the monetary contributions of the human, to participate in the game, then where is the line drawn between the representational rights of the avatar as a virtual citizen, and the human as a real citizen in a virtual world? T. L. Taylor observes that

Outside of any individual player’s time the account is in fact devoid of meaning. It takes a player to create a character and it takes the time of the player to develop that character. Through their labor they imbue it with qualities, status, accomplishments. Indeed, while the owners of a game provide the raw materials through which users can participate in a space, it is in large part only through the labor of the players that dynamic identities and characters are created, that culture and community come to grow.⁶

As further cases of the kind illustrated above, continue to arise and blur the lines of virtual citizenship, an onus is placed on Canadian lawmakers, and indeed lawmakers abroad, to determine exactly where the rights of the virtual citizen reside. Further legal investigation, possibly leading to a ruling on the exact nature of virtual worlds is necessary to establish legal precedent for the rights of Canadian citizens’ dual citizenships in the worlds of Azeroth, Norrath, Second Life, and the plethora of other worlds emerging daily.

Blurred Labour and Property Rights

Unlike common law property – say shoes, intellectual property in virtual worlds, computer games, and other code-based intellectual property are “non-rival goods”. So for example, when I am wearing my shoes, you are not. In a computer game or virtual world however, we could both be wearing my shoes – or at least my player’s shoes. Non-rival goods notwithstanding, some distinctions between intellectual property and common law property still apply.

Right now there is a man somewhere entering a shoe store. Inside the store, television monitors are broadcasting information on the latest footwear. Other customers are browsing through the displays piled high with shoeboxes. The man speaks to the salesperson and then selects a pair of sneakers. As he leaves the store he realizes the shoes are not exactly what he saw advertised on the television screens. Confused, the man backtracks into the store and asks to make a return. “We don’t give refunds here,” replies the salesperson, “this isn’t real life”.⁷

¹ M.Ed. Games Designer, Video Games Researcher, Ph.D. Student at Memorial University.

² M.A., Ph.D. (Ed. Tech.), LL.M. (Tech. Law). Professor at Memorial University.

³ World of Warcraft, <http://www.worldofwarcraft.com>

⁴ Fairfield, J. (July 2007). Anti-Social Contracts: The Contractual Governance of Online Communities. Washington & Lee Legal Studies Paper No. 2007-20; Indiana Legal Studies Research Paper No. 89, at SSRN: <http://ssrn.com/abstract=1002997>

⁵ Dougherty, Candidus, E-Commerce Law Report: MDY Industries, LLC Vs. Blizzard Entertainment, Inc. (August 18, 2008). E-Commerce Law Reports, Vol. 8, No. 2, August 2008. Available at SSRN: <http://ssrn.com/abstract=1235602>

⁶ Taylor, T.L. 2002. Whose game is this anyway: Negotiating corporate ownership in a virtual world. Proceedings of Computer Games and Digital Culture Conference, ed. Frans Mayra. Tampere: Tampere University Press, p 232.

⁷ Passman, M. (2009). Transactions of virtual items in virtual worlds, Albany Law Journal of Science and Technology, p261

Research article

Open Access

European citizens' use of E-health services: A study of seven countries

Hege K Andreassen*¹, Maria M Bujnowska-Fedak², Catherine E Chronaki³, Roxana C Dumitru⁴, Iveta Pudule⁵, Silvina Santana⁶, Henning Voss⁷ and Rolf Wynn^{1,8}

Address: ¹Norwegian Centre for Telemedicine, University Hospital of Northern Norway, Tromsø, Norway, ²Department of Family Medicine, Wrocław Medical University, Wrocław, Poland, ³Foundation for Research and Technology, Heraklion, Greece, ⁴Chair of Medical Informatics, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany, ⁵Health Promotion State Agency, Riga, Latvia, ⁶Universidade de Aveiro, Aveiro, Portugal, ⁷Danish Centre for Health Telematics, Odense, Denmark and ⁸Department of Clinical Psychiatry, University of Tromsø, Tromsø, Norway

Email: Hege K Andreassen* - hege.andreassen@telem.no; Maria M Bujnowska-Fedak - mbujnowska@poczta.onet.pl; Catherine E Chronaki - chronaki@ics.forth.gr; Roxana C Dumitru - Corina.Dumitru@imi.med.uni-erlangen.de; Iveta Pudule - iveta.pudule@vvva.gov.lv; Silvina Santana - silvina@egi.ua.pt; Henning Voss - HVO@health-telematics.dk; Rolf Wynn - rolf.wynn@gmail.com

* Corresponding author

Published: 10 April 2007

Received: 3 July 2006

BMC Public Health 2007, 7:53 doi:10.1186/1471-2458-7-53

Accepted: 10 April 2007

This article is available from: <http://www.biomedcentral.com/1471-2458/7/53>

© 2007 Andreassen et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: European citizens are increasingly being offered Internet health services. This study investigated patterns of health-related Internet use, its consequences, and citizens' expectations about their doctors' provision of e-health services.

Methods: Representative samples were obtained from the general populations in Norway, Denmark, Germany, Greece, Poland, Portugal and Latvia. The total sample consisted of 7934 respondents. Interviews were conducted by telephone.

Results: 44 % of the total sample, 71 % of the Internet users, had used the Internet for health purposes. Factors that positively affected the use of Internet for health purposes were youth, higher education, white-collar or no paid job, visits to the GP during the past year, long-term illness or disabilities, and a subjective assessment of one's own health as good. Women were the most active health users among those who were online. One in four of the respondents used the Internet to prepare for or follow up doctors' appointments. Feeling reassured after using the Internet for health purposes was twice as common as experiencing anxieties. When choosing a new doctor, more than a third of the sample rated the provision of e-health services as important.

Conclusion: The users of Internet health services differ from the general population when it comes to health and demographic variables. The most common way to use the Internet in health matters is to read information, second comes using the net to decide whether to see a doctor and to prepare for and follow up on doctors' appointments. Hence, health-related use of the Internet does affect patients' use of other health services, but it would appear to supplement rather than to replace other health services.

Background

There is a considerable demand for health-related information in the population, and the Internet is gaining ground as a central source of such information [1,2]. In the US, studies have found that between 56 % and 79 % of Internet users seek health information [3-6]. Starting out as a grassroots phenomenon much used by individual patients operating on their own and often offered by idealists as well as by commercial interests, Internet health sites and other electronic communication tools targeting patients are now important policy instruments for both public and private health providers. In recent years, we have seen national health authorities beginning to focus on e-health services such as electronic health cards, electronic patient records and health portals, including the English NHS Direct Online, the German Telematic Platform, and the Danish Sundhed.dk. In the medical community, expectations about the Internet are mixed. On one hand, the Internet has been described as having the potential to empower patients and stimulate patient participation [7-10]. On the other hand, potential dangers such as the dissemination of inaccurate information and inappropriate use have been stressed [11-13]. Earlier European studies have shown that the use of the Internet for health purposes varied in different parts of Europe [14,15]. As the dissemination of e-health services is growing along with general Internet use, there is a need to improve our knowledge on how these services are used, by whom and with what consequences. Two research questions were pursued in the present study; Do the users of Internet health services differ from the general population with respect to health and demographic variables? And, which health related Internet activities are most common? Further, we investigated citizens' expectations concerning the provision of e-health services by doctors.

Methods

A study group of 20 researchers designed a questionnaire for computer-assisted telephone interviews (CATI). The questionnaire was piloted with 100 individuals to ensure the comprehensibility of the wording and internal validity. It was designed in English and translated into the other languages by means of the dual focus approach [16]. This approach differs from the translation-back translation method in that it focuses on conceptual equivalence as well as on equivalence in wording and grammar. The aim is to reduce potential cultural bias in the questionnaire. The survey was conducted during the period October to November 2005. Random digit dialling in stratas ensured a randomised representative sample of the populations (age group 15 – 80 years) of seven European countries. The telephone penetration was estimated to be close to 100 % in Norway, Denmark, and Germany. In Poland it was estimated to be 63 %, in Latvia 93 %, in Greece 87 %, and in Portugal 65 %. Mobile phone numbers were

included in Norway, Denmark, Germany, and Latvia. Sampling continued until we had approximately 1000 completed interviews from all countries, except Portugal where 2000 interviews were conducted as health-related Internet use was expected to be low. Calculating a response rate is difficult when this sampling procedure is used, as a required number of responses is set before sampling starts, and sampling actually continues until the required number is obtained. The polling agencies conducting the interviews were instructed to follow standard procedures relating to contacting a replacement if a person originally selected for interview was unavailable (i.e. because of incorrect phone number, not answering the phone, not at home, or unwilling to participate). Nevertheless, we lack accurate data from all agencies relating to the number of people who were contacted in order to achieve the final number of completed interviews. A population weight was used to correct for differences in the sizes of the countries' populations for total estimates and logistic regression. No variables had more than 5% missing data. National ethics committees from all countries were informed and had no objections to the survey. We analysed the data by performing descriptive statistics and logistic regression analysis. SPSS version 12.0 was used for all analyses.

Results

The total sample consisted of 7934 respondents; out of these 4714 reported that they were Internet users. After weighting for population size, we had a total sample of 7903, of which 4906 were Internet users.

Before weighting, we calculated the proportion of Internet health users in each country (Table 1). Health-related use of the Internet was most frequent in the Northern countries, with Denmark (62 %), and Norway (59 %) topping the list, followed by Germany (49 %). The Eastern countries, Poland and Latvia, reported 42 % and 35 % health-related use of the Internet respectively, while the Southern countries had the lowest proportion of Internet health users with 30 % in Portugal and 23 % in Greece. In the sub-sample of Internet users, the differences between the countries were smaller, but a chi-square test showed that the differences between the Northern (74 % Internet health users), East-European (72 %) and Southern countries (60 %) were significant ($\chi^2_{(2,4714)} = 88, 5, p < 0.001$), despite the high score in Poland (79 %).

In the joint population of the seven countries, a total of 44 % (71 % Internet users) reported having used the Internet for health purposes (Table 2). In the general population, men were the most active health users on the Internet (47 % men, 42 % women). However, when Internet access was controlled for and we concentrated on those who were online, women tended to use the Internet more for

Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination

Anna Gagliardi, Alejandro R Jadad

Abstract

Objective This study updates work published in 1998, which found that of 47 rating instruments appearing on websites offering health information, 14 described how they were developed, five provided instructions for use, and none reported the interobserver reliability and construct validity of the measurements. **Design** All rating instrument sites noted in the original study were visited to ascertain whether they were still operating. New rating instruments were identified by duplicating and enhancing the comprehensive search of the internet and the medical and information science literature used in the previous study. Eligible instruments were evaluated as in the original study. **Results** 98 instruments used to assess the quality of websites in the past five years were identified. Many of the rating instruments identified in the original study were no longer available. Of 51 newly identified rating instruments, only five provided some information by which they could be evaluated. As with the six sites identified in the original study that remained available, none of these five instruments seemed to have been validated. **Conclusions** Many incompletely developed rating instruments continue to appear on websites providing health information, even when the organisations that gave rise to those instruments no longer exist. Many researchers, organisations, and website developers are exploring alternative ways of helping people to find and use high quality information available on the internet. Whether they are needed or sustainable and whether they make a difference remain to be shown.

Introduction

The quality of health information on the internet became a subject of interest to healthcare professionals, information specialists, and consumers of health care in the mid-1990s. Along with the rapid growth of healthcare websites came a number of initiatives, both academic and commercial, that generated criteria by which to ensure, judge, or denote the quality of websites offering health information. Some of these rating instruments took the form of logos resembling "awards" or "seals of approval" and appeared

prominently on the websites on which they were bestowed.

In 1997 we undertook a review of "award-like" internet rating instruments in an effort to assess their reliability and validity.¹ We hypothesised that if the rating instruments were flawed they might influence healthcare providers or consumers relying on them as indicators of accurate information. Instruments were eligible for review if they had been used at least once to categorise a website offering health information and revealed the rating criteria by which they did so. The rating instruments were evaluated according to, firstly, a system for judging the rigour of the development of tools to assess the quality of randomised controlled trials² and, secondly, whether their criteria included three indicators suggested as appropriate for judging the quality of website content:³⁻⁵ These indicators were authorship (information about authors and their contributions, affiliations, and relevant credentials), attribution (listing of references or sources of content), and disclosure (a description of website ownership, sponsorship, underwriting, commercial funding arrangements, or potential conflicts of interest). These criteria were selected for use in the original study because they could be rated objectively.

Our original study found that of 47 rating instruments identified, 14 described how they were developed, five provided instructions for use, and none reported the interobserver reliability and construct validity of the measurements. The review showed that many incompletely developed instruments were being used to evaluate or draw attention to health information on the internet.

The purpose of this study is to update the previous review of award-like rating instruments for the evaluation of websites providing health information and to describe any changes that may have taken place in the development of websites offering health information to practitioners and consumers with respect to the quality of their content.

Methods

We visited the websites describing each of the rating instruments noted in the original study to ascertain whether they were still operating. If internet service was disrupted for technical reasons or if sites were not

See Education and debate p 606

Graduate Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
Anna Gagliardi
senior research associate

Departments of Health Policy, Management and Evaluation, and Anaesthesia, University Health Network, University of Toronto, Toronto, ON, Canada

Alejandro R Jadad
professor

Correspondence to:
A R Jadad, Director, Centre for Global eHealth Innovation, University Health Network, Toronto General Hospital, Fraser Elliott Building, 4th Floor, 190 Elizabeth Street, Toronto, ON M5G 2C4, Canada
ajadad@uhmres.utoronto.ca

BMJ 2002;324:569-73

Papers

Table 1 Summary of criteria for rating instruments

Rating system	Health specific scope	Silberg et al ⁶ and Wyatt ⁷			Moher et al ⁸				Criteria changed from original study	
		Authorship	Attribution	Disclosure	Type of instrument	No of Items	Scale development	Reliability		Instructions
Previously reviewed sites										
American Medical Association (ama-assn.org/ama/pub/category/13952.html)	Y	✓	✓	✓	U	8	NR	NR	NR	N
Argus Clearinghouse seal of approval (clearinghouse.net/ratings.html)	Y	NC	NC	NC	U	U	NR	NR	NR	N
GrowthHouse excellence award (growthhouse.org/award.html)	Y	NC	U	U	S (stars)	U	NR	NR	U	N
Health on the Net Foundation code of conduct (www.hon.ch/HONcode/Conduct.html)	Y	✓	✓	✓	Logo	8	U	NR	Y	N
Medaille d'Or for website excellence (arschuld.co.uk/award/select.html)	N	NC	NC	NC	S (medals)	U	NR	NR	U	N
Oncolink's editors' choice awards (oncolink.upenn.edu/ed_choice/)	Y	NR	NR	NR	Logo	NR	NR	NR	NR	Y
Newly identified sites										
World wide web health awards (healthawards.com/www/ha/s2001Webawards/assessment.htm)	Y	NC	NC	NC	Logo	U	NR	NR	NR	—
HardinMD clean bill of health (hb.uswa.edu/hardin/md.cbh.html)	Y	NC	NC	NC	Logo	U	NR	NR	NR	—
Nutrition Navigator among the best (navigator.tufts.edu)	Y	NC	✓	NC	S	5	NR	NR	U	—
Pacific Bell knowledge network blue web'n (kn.pacbell.com/wired/blueWebn/rubric.html)	N	✓	✓	U	S	U	NR	NR	NR	—
Health Improvement Institute Aesculapius award for rating sites (hii.org)	Y	✓	✓	✓	Logo	U	NR	NR	NR	—

Y=yes; N=no; S=scale; U=unclear; NR=not reported; NC=not considered; —=considered.

available on first visit, we attempted a connection on one further occasion.

The search strategies, inclusion and exclusion criteria, and techniques for data extraction were similar to those used in the original review.¹ We used the following sources to identify new rating instruments:

- A search to 7 September 2001 of Medline, CINAHL, and HealthSTAR (from December 1997) using [(top or rat: or rank: or best) and (internet or web) and (quality or reliab: or valid:)]
- A search of the databases Information Science Abstracts, Library and Information Science Abstracts (1995 to September 2001), and Library Literature (1996 to September 2001) using [(rat: or rank: or top or best) and (internet or web or site) and (health:)]
- A search to September 2001 using the search engines Lycos (lycos.com), Excite (excite.com), Yahoo (yahoo.com), HotBot (hotbot.com), Infoseek (go.com), Looksmart (looksmart.com), and Google (google.com) with [(rate or rank or top or best) and (health)]. Open Text (opentext.com) and Magellan (magellan.com), which were used in the first study, no longer function as internet search engines
- A review of messages about rating instruments and the quality of health related websites posted to the Medical Library Association listserv medlib-l (listserv.acsu.buffalo.edu/archives/medlib-l.html) and the Canadian Health Libraries Association listserv canmedlib-l (lists.mun.ca/archives/canmedlib.html)
- A search of the American Medical Informatics Association's 1998, 1999, 2000, and 2001 annual symposium programmes (wwwamia.org) for mention of health information on the internet
- A search of the *Journal of Medical Internet Research* (September 1999 to September 2001) for mention of

evaluations of the quality of health information on the internet (www.jmir.org)

• A search of the online archive of the magazine *Internet World* (www.internetworld.com) (January 2000 to September 2001) for mention of health information on the internet.

We also reviewed relevant articles referenced in identified studies and links available on identified websites. We did not search the discussion list Public Communication of Science and Technology, which was consulted in the original study.

We stopped searching for rating instruments on 22 September 2001. Rating instruments were eligible for inclusion in the review if it was possible to link from their award-like symbol to an available website describing the criteria used by an individual or organisation to judge the quality of websites on which the award was bestowed. We excluded rating instruments from review if they were used only to rate sites offering non-health information or did not provide any description of their rating criteria. In contrast to the initial study, we did not contact the developers of rating instruments to request information about their criteria if it was not publicly available on their website.

We identified the website, group, or organisation that developed each eligible rating instrument, along with its web address. The two authors independently evaluated each rating instrument according to its validity (number of items in the instrument, availability of rating instructions, information on the development of rating criteria, and evaluation of interobserver reliability) and incorporation of the proposed criteria for evaluation of internet sites: authorship, attribution, and disclosure.²⁻⁴

Marketing Strategy and the Internet: An Organizing Framework

P. Rajan Varadarajan
Manjit S. Yadav

Texas A&M University

Competitive strategy is primarily concerned with how a business should deploy resources at its disposal to achieve and maintain defensible competitive positional advantages in the marketplace. Competitive marketing strategy focuses on how a business should deploy marketing resources at its disposal to facilitate the achievement and maintenance of competitive positional advantages in the marketplace. In a growing number of product-markets, the competitive landscape has evolved from a predominantly physical marketplace to one encompassing both the physical and the electronic marketplace. This article presents a conceptual framework delineating the drivers and outcomes of marketing strategy in the context of competing in this broader, evolving marketplace. The proposed framework provides insights into changes in the nature and scope of marketing strategy; specific industry, product, buyer, and buying environment characteristics; and the unique skills and resources of the firm that assume added relevance in the context of competing in the evolving marketplace.

In a growing number of product-markets, the competitive landscape has evolved from a predominantly physical marketplace to a broader marketplace encompassing both the physical and the electronic marketplace. The emergence of the electronic marketplace has been associated with a number of developments, including the following:

- Greater information richness of the transactional and relational environment

Journal of the Academy of Marketing Science,
Volume 30, No. 4, pages 296-312.
DOI: 10.1177/009207002236907
Copyright © 2002 by Academy of Marketing Science.

- Lower information search costs for buyers
- Diminished information asymmetry between sellers and buyers
- Electronic spatial proximity of buyers and sellers
- Greater temporal separation between time of purchase and time of possession of physical products purchased in the electronic marketplace
- Greater temporal proximity between time of purchase and time of possession of digital products purchased in the electronic marketplace

Paralleling these developments, the business literature has witnessed a number of new additions to its lexicon (e.g., e-business, e-commerce, e-procurement, e-services, e-customer relationship management, e-alliance partner relationship management, e-supplier relationship management, e-supply chain management, etc.). Concurrently, scholarly research in marketing has focused on such issues as new models of communication (Hoffman and Novak 1996), buyers' and sellers' incentives to participate in the electronic marketplace (Alba et al. 1997; Burke 1997; Grewal, Comer, and Mehta 2001), and the migration of products to the electronic marketplace (Peterson, Balasubramanian, and Bronnenberg 1997; Yadav and Varadarajan 2001). The contextual relevance of factors such as industry structure characteristics (e.g., market dispersion and market thinness), product characteristics (e.g., product digitizability), and buying environment (e.g., information search costs) for competing in the electronic marketplace have also been explored in recent research (Alba et al. 1997; Bakos 1991, 1997; Balasubramanian, Krishnan, and Sawhney 2000; Benjamin and Wigand 1995; Blattberg and Deighton 1991; Burke 1996; Gurbaxani and Whang 1991; Hoffman and Novak 1996, 1997; Lynch and Ariely 2000; Rayport and Sviokla 1994). Contributions from strategy (e.g., Porter 2001), manage-

ment information systems (e.g., Adam and Yesha 1996; Huber 1990; Malone 1997; Malone, Yates, and Benjamin 1989; Weill and Broadbent 1998), economics (e.g., Bakos 1991, 1997; Brynjolfsson and Seidmann 1997; Gurbaxani and Whang 1991; Shapiro and Varian 1999), and technology (e.g., Clark 1996; Oliver 1996) complement the above research streams in marketing.

Against this backdrop, this article focuses on the implications for marketing strategy of a competitive landscape that is evolving from a physical marketplace to one encompassing both the physical and the electronic marketplace. More specifically, our objectives are to

- delineate the nature and scope of competitive marketing strategy in reference to competing in the physical and the electronic marketplace; and
- develop a conceptual framework that delineates representative industry structure, firm, product, buyer, and buying environment characteristics pertinent to competing in the physical and the electronic marketplace.

The remainder of this article is organized as follows. First, we provide a brief overview of the evolving electronic marketplace and extant perspectives on competitive strategy. Second, building on extant literature, we present a conceptual framework focusing on selected drivers and outcomes of competitive strategy. Third, we present a refinement and extension of the general framework by delineating certain *additional* industry structure, firm, product, buyer, and buying environment characteristics that are particularly pertinent to competing in an electronic marketplace (and, by extension, in the new competitive landscape that encompasses both the physical and the electronic marketplaces). The proposed framework focuses on new challenges and opportunities for ongoing producer firms (legacy businesses) as they seek to understand the implications of the evolving marketplace for competitive marketing strategy.

THE ELECTRONIC MARKETPLACE: CONCEPTUALIZATION AND EVOLUTION

Conceptualization of the Electronic Marketplace

The Internet, a rapidly expanding global computer and communications infrastructure, has facilitated the emergence of computer-mediated environments that serve as electronic marketplaces for buyers and sellers. Bakos (1991) defined the *electronic marketplace* as "an inter-organizational information system that allows participating buyers and sellers to exchange information about prices and product offerings" (p. 296). According to Barrett and

Konsynski (1982), interorganizational information systems span organizational boundaries, linking firms to their customers and/or suppliers. This view of the electronic marketplace, while useful, appears restrictive in two respects. First, the term *interorganizational information system* seems to suggest an infrastructure created by and for organizations. In reality, the role played by consumer-controlled access devices (e.g., personal computers and personal digital assistants) continues to increase. Therefore, in conceptualizing the electronic marketplace, the broader term *networked information system* seems more appropriate than the seemingly restrictive term *interorganizational information systems*. Second, the nature and scope of activities that occur in the electronic marketplace often extend well beyond the "exchange of information about prices and product offerings." In fact, what often transpires after the exchange of information—a transaction and other activities related to the transaction—represent important functions that are also facilitated by the electronic marketplace and should, therefore, be noted. While the specific implementation of such functions may vary across firms, any conceptualization of the electronic marketplace must more fully encompass the scope of activities that may occur in such a setting. Accordingly, we conceptualize the electronic marketplace as a *networked information system that serves as an enabling infrastructure for buyers and sellers to exchange information, transact, and perform other activities related to the transaction before, during, and after the transaction*. A brief elaboration of this conceptualization follows.

As can be noted, an *electronic marketplace* performs essentially the same set of functions as a *physical marketplace*—both bring buyers and sellers together. While they both share this important common purpose, electronic and physical marketplaces do have certain distinguishing characteristics. The most obvious salient difference, of course, is that the enabling infrastructure is electronic rather than physical. This difference, in turn, leads to a number of other differences that are worth mentioning. For instance, variations such as the following are increasingly prevalent in the electronic marketplace: private electronic marketplaces that serve either multiple buyers and a seller (e.g., buyers purchasing tickets directly from an airline's Web site) or multiple sellers and a buyer (e.g., sellers participating in a live reverse auction at a prespecified date and time in response to a call for bids posted by a buyer). Such variations can occur in the physical marketplace as well, but they can be scaled much more readily in the context of the electronic marketplace. A second distinguishing characteristic pertains to the increasingly prominent role played by *other participants* in the electronic marketplace—entities other than buyers and sellers who provide value-added services to buyers and/or sellers. For example, interacting with other marketplace participants may involve activities such as a buyer accessing information

What Attracts Customers to Online Stores, and What Keeps Them Coming Back?

David J. Reibstein

University of Pennsylvania

Many businesses on the Internet in the late 1990s spent wildly, doing whatever it might take to attract customers to their sites. It soon became clear that the challenge was not simply to bring the customers in the door but also to retain these customers for future purchases. The quest was on to discover what tactics had the most appeal to Internet shoppers. This study reveals survey and behavioral data drawn from Internet customers that reflect what was most important to the Internet shoppers and compare the factors for attraction versus retention. Since many have viewed the Internet as creating more perfect information for the buyer, the question arises as to how important price will be in the purchase process. What becomes clear from the analysis is that what attracts customers to the site are not the same dimensions critical in retaining customers on a longer term basis.

As the Internet unfolded and the world of dot.com businesses began to emerge, it became clear that the race was on to acquire customers. Many dot.com businesses were under the belief that the first to gain customers would be the "winners" in this space. As such, in the late 1990s, we witnessed excessive amounts of spending on advertising, such as Super Bowl 1999 and 2000 advertising by dot.coms, numerous promotions and "give-aways," and other forms of enticement to attract customers to Internet sites. Much attention was given to the conviction of "first-mover advantages" on the Internet. As a result, there was

Journal of the Academy of Marketing Science,
Volume 30, No. 4, pages 465-473.
DOI: 10.1177/0092072002236918
Copyright © 2002 by Academy of Marketing Science.

little time to study what worked online and what did not. There was an urge to be there first.

Of course, the premise of the first-mover advantage on the Internet was the belief that once the customers came, they would be "locked in" to using that site, thereby making it more difficult for later entrants to be able to attract customers. Hence, the quest was on for customer acquisition.

Much of the willingness to wantonly spend in order to acquire customers was based on some either explicit or implicit notion of the lifetime value of the customer, a much-developed concept in the marketing literature. Why else would anyone be willing to spend more to acquire a customer than the margins generated from the one-time customer sale? This became all the more apparent when the next rallying theme on the Internet was that of customer retention. The central focus moved internally to assess what operational features were most effective in keeping customers so they would continue to shop at a particular site.

The Internet, while offering promise to the numerous aspiring entrepreneurs, is a dangerous territory. Because of the ease of switching and the ability to quickly gather near-perfect information, customers are awarded with a new set of power tools in their decision making. Armed with the latest information and prices of what a product is being sold for at numerous sites, there is little to inhibit customers from switching suppliers or from changing where they would shop. This makes the retention task all the more difficult.

Without doubt, much of the expenditure for both the acquisition and the retention objectives was based on intuition and what the general managers felt were the most likely methods to succeed. Undoubtedly, much of the expenditure was also stimulated by managers who

emulated what others had done, believing that these are the necessary steps in acquiring and retaining customers. Hence, we were able to observe spiraling marketing expenditures.

The purpose of this article is to explore from the customers' perspective what attracts them to a specific site and which characteristics of a site keep them coming back. In particular, we will explore the role of price both for the acquisition process and for customer retention. This should provide guidance to future e-tailors about the best methods for attracting and retaining customers and the importance of having the lowest price, based on market data. This article provides a background of the literature on decision making in the digital environment and how it might be different, a description of the data that are explored in this study, the results, and the implications for doing business online.

BACKGROUND

In a surprisingly short amount of time, a literature has built up on consumer decision making in the digital environment. Dholakia and Bagozzi (2001) did an excellent job of summarizing much of what has been written about how consumers make their purchasing decisions in the new digital environment. Similarly, Häubl and Trifts (2000) discussed consumer decision making and the impact of decision aids in the process. Both are excellent references.

One perspective is that the Internet will allow customers to become more efficient in their buying process. This efficiency results primarily because the cost of information, as well as the cost and time required of acquiring information, is so low (Bakos 1997). With information being exceedingly easy to access, the resulting outcome will be that consumers have the capability of becoming fully informed with regard to their options (Bakos 1997; Brynjolfsson and Smith 1999). Thus, they will be able to make better decisions with less required effort.

Often, the amount of information that can be delivered to the consumer is constrained by the size of the package and the space available on which to deliver a message, the time available in a 30-second spot, or the space available in a print ad. Alternatively, the cost of delivering the information could be quite high if one were to buy the ad space to provide full communication of information. Given that information requirements may differ across consumers, the total amount of possibly required information could be daunting and prohibitively expensive. In the digital environment, these constraints are much less severe. Package size is not a factor, nor is ad space. In the digital world, the information is not limited by the physical space (Alba et al. 1997; Johnson, Lohse, and Mandel 1997).

As pointed out by Dholakia and Bagozzi (2001), there is a downside to this information-abundant environment—that is, the possibility of information overload and the cognitive costs that might be involved (Bettman, Johnson, and Payne 1991; Malhotra 1982). The potential consequence of this excess of information is, perhaps, customers making suboptimal decisions and/or developing decision heuristics to simplify the decision task. But how does the decision process differ in an environment when information is plentiful and easy to access and perhaps there is more information than the customer can process?

The Role of Prices in the Choice Process

The Argument for Lower Prices Online

Many authors have speculated about the role of prices in the "efficient market" setting. Customers armed with full information of what prices are charged by different retailers, particularly for the same product, will be able to make the most economical decision. A barrier to finding the lowest price for an item in the "bricks-and-mortar" world has always been the lack of customer information and the cost of acquiring this information.

The process of searching for information online has been even more accelerated with the availability of electronic agents that automate the search process. In these cases, consumers can find all (or most) of the Web sites that sell a specified product and list the product's prices at these various sites. The role of consumer behavior when confronted with such electronic agents could easily be altered (Häubl and Trifts 2000; West et al. 1999). It is possible that with the introduction of electronic agents, the factors that attract and retain customers to a site might be different because of the ease of comparison across sites on certain dimensions.

Brynjolfsson and Smith (1999) found that prices online are 8 to 15 percent lower than the prices for comparable products in traditional retail outlets. This might be because of lower direct costs to supply the product (i.e., no rent, lower or centralized inventory, etc.). It might also be because of more price competition online—more competitors with more focus on price. A third potential reason for this might be the removal of the physical monopoly or an advantage any one retailer might have over another because of its proximity to the customer; that is, the customer has to incur an additional travel and time cost to go to another retail outlet. This cost to the customer does not exist to the same extent in the digital world, and thus the local monopoly is lessened.

Because of the visibility of prices as a comparison variable across e-tailors (the primary focus of most electronic agents), it puts added price pressure on each of the e-tailors. Last, the customer who is supplied with full information (or the "efficient consumer") may be able to make a more

Learning from Linked Open Data Usage: Patterns & Metrics

Knud Möller, Michael Hausenblas,
Richard Cyganiak, Siegfried Handschuh
Digital Enterprise Research Institute (DERI),
National University of Ireland, Galway
IDA Business Park, Lower Dangan, Galway,
Ireland

knud.moeller@deri.org,
michael.hausenblas@deri.org,
richard.cyganiak@deri.org,
siegfried.handschuh@deri.org

Gunnar Aastrand Grimnes
German Research Center for Artificial
Intelligence (DFKI) GmbH, Knowledge
Management Group
D-67663 Kaiserslautern, Germany
grimnes@dfki.uni-kl.de

ABSTRACT

Although the cloud of Linked Open Data has been growing continuously for several years, little is known about the particular features of linked data usage. Motivating why it is important to understand the usage of Linked Data, we describe typical linked data usage scenarios and contrast the so derived requirement with conventional server access analysis. Then, we report on usage patterns found through an in-depth analysis of access logs of four popular LOD datasets. Eventually, based on the usage patterns we found in the analysis, we propose metrics for assessing Linked Data usage from the human and the machine perspective, taking into account different agent types and resource representations.

Keywords

linked data, Web of Data, access, usage patterns

1. INTRODUCTION

Linked Open Data (LOD) [6] is a recent community effort to alleviate the problem of missing, sufficiently inter-linked datasets on the Web of Data. Through this effort, a significant number of large-scale datasets¹ have now been published in the LOD cloud², which is growing constantly.

At time of writing, 6.7 billion RDF triples and over 140 million links between datasets are available [5]. Though research is known to investigate search engine crawls and logs [8, 13, 17] the usage behaviour of agents — humans and machines alike — concerning linked data has so far not gained much attention. Hence, in this work we analyse access logs of LOD data sets in order to *learn how Linked Data is used*. Our contribution is twofold: (i) We report on usage patterns found in LOD datasets. (ii) Based on our observa-

¹For example, DBpedia (<http://dbpedia.org/>), BBC music (<http://www.bbc.co.uk/music/>), LinkedGeoData (<http://linkedgeodata.org/>), and only recently by the New York Times (<http://data.nytimes.com/>)

²http://www4.wiwiw.fu-berlin.de/bizer/pub/lod-datasets_2009-07-14.html

tion we propose an initial set of dedicated Linked Data usage metrics. As a starting-point we briefly review related work (Sect. 1.1) and discuss the challenges of analysing linked data usage in Sect. 2. We then report on the results of our analysis in Sect. 3 and propose usage metrics for Linked Data in Sect. 4. Sect. 5 concludes our work and sketches next steps.

1.1 Related Work

Analysing server logs is as old as the Web itself [22]. To this end, research has focused typically [27, 21, 26] on: (i) server-side: performance, optimisation (load balancing, etc.) (ii) client-side: customisation, etc. Our work can be seen as a case of Web use mining [17, 26] in the wider sense, with a focus on the analysis of semantically-enabled Web sites [20]. However, to the best of our knowledge none of the existing work has looked specifically at LOD or SPARQL (the query language of the Web of linked data) usage.

2. MOTIVATION

Much as the current Web (of documents) is mainly targeting human users, a particular strength of linked data is that applications can use it directly [12]. We hence differentiate two fundamental types of consumers concerning the usage of LOD: (i) **Human Users**, equipped with a generic Linked Data browser [14] such as Tabulator [4] or Sigma [7] on the one hand, and (ii) **Machine Agents**, that is, a piece of software not under the direct control of a human, on the other hand. One would assume that human and machine agents differ in terms of usage patterns. Whereas we suspect human users to browse the LOD datasets in a more traditional, rather unpredictable sense, we would imagine machine agents to be more “focused”: machine agents are typically based on a fixed rule set encoded in their program. Additionally, if LOD is primarily targeting applications rather than humans, we would expect the majority of the usage caused by machine agents.

2.1 Motivating Challenges

Our motivation to better understand the usage of LOD data is tightly related to machine agents. There are a couple of challenges concerning LOD usage, especially from the machine agents point of view, which have so far been neglected by and large:

- Concerning **reliability**: with the recent additions from the commercial domain such as BBC and NY times, the LOD cloud developed into a commercial-strength global database. If one is about to use LOD from an application, the availability of the data is crucial.
- Concerning the **peak-load**: certain LOD datasets (or certain entities in LOD datasets) may be requested more frequently than others; this might be due to events (such as conferences, celebrations, launches, etc.) or due to their role as linking hubs.
- Concerning **performance**: knowing what queries are being executed against your store helps configuring caches and indexes.
- Concerning **usefulness**: what parts of your data is being accessed often, what links are people finding and following, or alternatively, searching for and NOT finding.
- Concerning **attacks**: with the growing commercial usage, targeted attacks, such as or spam³ have to be taken seriously.

In order to address the above challenges, one needs to understand in-depth who is using the data and how it is used. To better understand the usage, we have performed an analysis of the server access logs of major LOD sites and report on the findings in the next section.

3. ANALYSIS OF LINKED DATA ACCESS LOGS

In this section, we are first going to give an overview of the four different evaluated datasets, and the source data available for each. Afterwards, we will look into a number of questions which aim to increase our understanding of linked data usage.

3.1 Source Data

In order to analyse the usage of linked data sites, answer questions relating to usage patterns and classify them according to the metrics proposed in this paper, we take the basic approach of examining the server log files of the sites in question. Such log files record each individual HTTP request that is made to a server, keeping information about such things as the requested URI, the HTTP method used, the origin of the request, the exact time of the request, the agent performing the request and details about the response of the server (for an example see Fig. 1). While different Web servers use different log formats, the scope of the data recorded in each format is similar. By far the most common format is the *common log format* (CLF)⁴, or the slightly extended *combined log format*. The latter was used by the servers hosting all four datasets in our analysis.

3.2 Evaluated Datasets

For our analysis, we had access to server log files from four different LOD sites: DBpedia, DBTune, RKBExplorer and

³<http://iandavis.com/blog/2009/09/linked-data-spam-vectors>

⁴CLF is an informal standard with no official reference. As a general introduction, we point the reader to http://en.wikipedia.org/wiki/Common_Log_Format.

SWC (aka “Semantic Web Dog Food”). All four sites differ greatly with respect to several of their basic characteristics, such as size (in number of RDF triples), connectedness in the LOD cloud, functionality beyond serving of linked data, etc. All four datasets together provide us with good coverage of the different types of datasets which make up the Web of linked data.

Below, we will give a brief introduction to each site, as well as an overview of some of their basic statistics in Tab. 1, such as size, period of time observed and number of hits in different categories. Specifically, we distinguish between requests to the SPARQL endpoints of each site and three related kinds of URIs which all reflect the same resource, in the sense that the *plain* resource URI is the identifier of a non-information resource [15] such as “WWW2009”, while the related *RDF* and *HTML* document URI are identifiers for information resources, or representations in different formats about WWW2009. This is discussed in more detail in Sect. 3.3.1. For orientation purposes, the total number of hits to a site is also given. Because we had access to different amounts of log file data for each site, Tab. 1 gives both the absolute numbers for each site, as well as the average per day. For reasons discussed below, some of the statistics could not be generated for the RKBExplorer data set.

3.2.1 Semantic Web Dog Food

The smallest dataset in our analysis in terms of RDF triples (~80,000 RDF) is served through the Semantic Web Conference metadata site (SWC or “Dog Food”) [19, 18]. SWC holds RDF data about a number of large, international conferences in the Web and Semantic Web area, such as WWW, ISWC and ESWC, as well as a growing number of workshops. For each such event, detailed data about papers, authors, events and other entities is provided, both as RDF and as HTML documents. For this dataset, we had the best coverage over time, comprising of log files spanning well over 1 1/2 years, from 21/07/2008–03/10/2010.

3.2.2 DBpedia

By far the largest dataset in our analysis is the DBpedia [2], which provides linked data based on an extraction of structured data from Wikipedia. Because of its wide coverage in background knowledge entities such as people, places, species, etc., DBpedia can be considered a hub within the Web of linked data, in that it is used as a point of reference by many other datasets. The DBpedia site serves both RDF and HTML documents about its resources. For DBpedia, we had access to server log files dating from 30/06/2009–25/10/2009 (i.e., 118 days).

3.2.3 DBTune

DBTune⁵ [24] is a meta-site which hosts different (currently 10) sub-datasets of linked data for a number of music-related non-LOD datasets, such as MusicBrainz, MySpace or Jamendo. While all datasets are hosted within the DBTune namespace, each of them is served in a slightly different manner, which makes an integrated analysis complicated. E.g., the MusicBrainz dataset is hosted through a D2R server instance and provides both HTML and RDF representations for its resources, for MySpace only RDF descriptions are provided at document-type URIs, while for Jamendo only RDF descriptions via SPARQL DESCRIBE

⁵<http://dbtune.org>

Vizster: Visualizing Online Social Networks

Jeffrey Heer
Computer Science Division
University of California, Berkeley

danah boyd
School of Information Management and Systems
University of California, Berkeley

ABSTRACT

Recent years have witnessed the dramatic popularity of online social networking services, in which millions of members publicly articulate mutual "friendship" relations. Guided by ethnographic research of these online communities, we have designed and implemented a visualization system for playful end-user exploration and navigation of large-scale online social networks. Our design builds upon familiar node-link network layouts to contribute customized techniques for exploring connectivity in large graph structures, supporting visual search and analysis, and automatically identifying and visualizing community structures. Both public installation and controlled studies of the system provide evidence of the system's usability, capacity for facilitating discovery, and potential for fun and engaged social activity.

CR Categories and Subject Descriptors: H.5 [User Interfaces] Graphical User Interfaces, I.3 [Methodology and Techniques] Interaction Techniques, K.8 [Personal Computing]

Keywords: social networks, visualization, graphs, community, data mining, exploration, play

1 INTRODUCTION

The advent of the Internet has given rise to many forms of online sociality, including e-mail, Usenet, instant messaging, blogging, and online dating services. In 2003, another form of online community acquired stunning popularity: online social networking services. In addition to descriptive personal profiles, members of such communities publicly articulate mutual "friendship" links with other members, creating a browseable network of social relations. Although pre-dated by other services such as sixdegrees.com in 1997, social networking services emerged as a veritable phenomenon with the dramatic rise of friendster.com, which rapidly amassed millions of users beginning in the spring of 2003. Friendster's success was closely followed by a number of other popular services, including Tribe.net and orkut.com, quickly earning these services the acronym "YASNS" or "Yet Another Social Networking Service."

Though the users of these systems have constructed massive graph structures of social connectivity, typical web interfaces to these social networks remain relatively impoverished, showing only the network connections of single individuals in a linear list on a web page. Articulated connections between one's own "friends" in these systems are obscured, and can be unearthed only by paging through each friend's profile page. Higher level patterns of community can be even harder to discern. This has problematic implications for members' ability to explore their online community and gauge both the scale and the individuals to which their self-reported personal information is exposed.

In this paper we present the design of Vizster, a visualization system for exploring such online social networks. Vizster builds

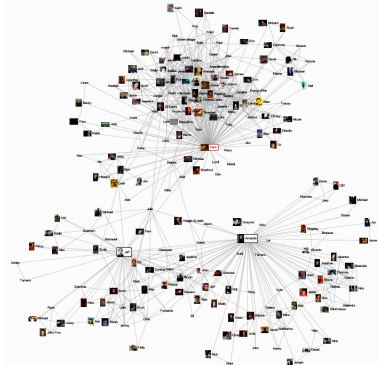


Figure 1: Vizster depicting three intersecting social networks

on ethnographic research of online social networking services and previous work in social network visualization to provide a system by which members of such online communities can explore their articulated social network in the playful manner that they desire. Our system is motivated by both social and technological concerns. On the social side, we attempted to better facilitate the discovery of people, connections, and communities to promote increased awareness of community structure and information exposure, while preserving (and hopefully enhancing) a fun and engaging online space. From a design perspective, this case study explores the mutually informing use of ethnographic techniques and visualization design to craft a domain-specific visualization system in a context as much characterized by play as by analysis.

Though the resulting system leverages existing techniques, these requirements resulted in a number of novel customizations and non-standard design decisions. Our design builds upon familiar node-link network layouts to provide novel techniques for visualizing graph inter-connectivity, supporting visual search and analysis of both network and profile data, customizing spring-embedded layouts, and automatically identifying and visualizing community structures. We have used the resulting system to visualize a 1.5 million member crawl of the popular Friendster service, collected in Winter 2003.

We first summarize related work in social network visualization and share relevant findings from an ethnographic investigation of the Friendster service. We then describe the design of the Vizster visualization in detail, followed by observations from a public installation and informal studies of the system's usage.

2 RELATED WORK

Visualization of social networks has a rich history, particularly within the social sciences, where node-link depictions of social

relations have been employed as an analytical tool since at least the 1930s. Linton Freeman documents the history of social network visualization within sociological research, providing examples of the ways in which spatial position, color, size, and shape can all be used to encode information [8]. For example, networks can be arranged on a map to represent the geographic distribution of a population. Alternatively, algorithmically generated layouts have useful spatial properties: a force-directed layout can be quite effective for spatially grouping connected communities, while a radial layout intuitively portrays network distances from a central actor. Color, size, and shape have been used to encode both topological and non-topological properties such as centrality, categorization, and gender.

In recent years, such approaches have been effectively used in the analysis of domains such as e-mail communication [7], early online social networks [1], and co-authorship networks in scientific publications [16]. There are a number of systems for generating such social network visualizations and performing statistical analyses for the purpose of sociological research, such as UCInet [20], JUNG [12], and GUESS [2].

In addition to sociological research, there have been numerous social visualization projects intended for end-users rather than outside researchers. Though often utilizing techniques seen in analytical domains, these systems present users with visualizations of their own online social world; they are non-anonymous and often perceived with a sense of social place [10] different from more detached analytical environments. Nardi *et al.*'s ContactMap [15] uses spatial grouping and color to redundantly code community groups within a visualization of a user's e-mail contacts. TouchGraph (<http://touchgraph.com>) uses a force-directed layout to present a network visualization of users of the LiveJournal online community, allowing personal networks to be expanded or contracted by user interaction. BuddyZoo (<http://buddyzoo.com>) analyzes users' instant messaging (IM) buddy lists to present a static network visualization of their IM contacts. Mutton's PieSpy [14] provides real-time, dynamic visualization of inferred social networks of Internet Relay Chat participants. Viégas *et al.*'s PostHistory and Social Network Fragments systems [21] visualize personal e-mail archives in both calendar and network views, including rich support for temporal filtering. They found these systems—especially their use in tandem—to be particularly well suited as a memory aid for past contexts and contacts.

Despite this wealth of social network visualization, we believe there is still a need for new designs and techniques, especially as articulated social networks become increasingly common in web services for signifying various kinds of relationship. Visualization of profile attributes unique to online social networks is needed, and techniques for incorporating analytical tools within the simplified domain of end-user visualization may prove useful.

3 FRIENDSTER DESCRIPTION

To better navigate the myriad design decisions we faced, we turned to an ethnographic study of the Friendster service. The following is the result of a 9-month participant-observation during 2003, including interviews, qualitative surveys and focus groups with over 200 Friendster early adopters [4, 6].

Friendster was designed to be an online dating site, complete with profiles, demographic and interest driven search, and a private messaging system. What made Friendster unique was its articulated social networking component and testimonial feature. Users were asked to declare "friends" on the system whose pictures would also appear on the profile when the friends confirmed the relationship. Friends could write testimonials that

would also appear on the profile. Both the friends and testimonials were intended to signal additional information about the person's character for those interested in dating the person. Yet, when the early adopters began to use the service, they did not view it as a dating service, but a site where they could gather and communicate with their friends, surf for entertaining profiles and explore public displays of identity and relationships [6].

Early adopters fell primarily into one of three identity groups: gay men, techies/bloggers, and Burners (people who attended the Burning Man art festival in Nevada every September). Each group had a different presentation aesthetic and their profiles conveyed the aspects of their identity most associated with the communities through which they joined the service. For example, Burners usually used their "Playa names" (the nickname they used at the festival), uploaded pictures of themselves at the festival and talked about their music tastes. Many early adopters had densely connected social groups, which aided Friendster's spread. As Friendster became more popular, the user population diversified. That said, the vast majority of users throughout 2003 were in their 20s, college-educated, politically liberal, and living in urban centers. Teenagers did join the service, reporting ages of 69 or 61 and 71 (16 and 17 inverted) because the site forbade users under 18. By marking their age this way, teen users were able to easily find each other. While the service began in the San Francisco Bay Area, its popularity quickly spread to Europe and to Singapore, Malaysia and the Philippines where it took off dramatically. New users primarily came through invitations from active users; thus, the user base represents the extended networks of early adopters.

When users joined the site, they were required to develop a profile that asked for demographic information (age, location, hometown) and a description of tastes (dating preferences, music, books, movies). Users typically created the profile based on the norms of the friends who invited them. As such, the network structure was also apparent in the profile norms of different groups. People regularly updated their profiles, adding new photos, changing tastes dependent on their friends and writing testimonials for their friends in order to encourage reciprocity.

Through ethnographic analysis, boyd [4] uncovered common social practices endemic to Friendster amongst early adopters. While dating did occur, most joined the service because all of their friends joined. Users began using the service by searching for friends and old friends, adding them to their network where appropriate. Because Friendster limited the visibility of profiles to four degrees (friends' friends' friends' friends), users were often motivated to connect to as many people as possible to expand their visibility. Friendster's design decision to list the most highly connected people on the front page further activated the competitive tendency to collect as many people as possible.

In addition to personal profiles, early adopters began creating fake profiles – "Fakesters" – to represent anything from famous media figures like Angelina Jolie to cities like New York, concepts like Love to everyday objects like Salt. Fakesters like Brown University helped connect people who attended that school while Fakesters like Homer Simpson helped connect fans of the TV character. The Fakesters increased the density of the network because they both helped friends find each other and connected new people through shared interests. The profiles created for Fakesters were often quite creative and motivated many users to surf for interesting profiles in a treasure-hunt fashion.

While most users surfed the network by clicking on the friends of other profiles, there was also a search tool available, which helped in finding friends by name and aided in finding potential dates or hookups. Users could also click on interests and find anyone with that interest marked.