Journal of the Association for Information Systems
SPECIAL ISSUE

Hurdles to Information Seeking: Spelling and Typographical Mistakes During Users' Online Behavior [1]

Eszter Hargittai
Department of Communication Studies
Institute for Policy Research
Northwestern University
hurdles05@webuse.org

Abstract

A refined approach to digital inequality requires that in addition to looking at differences in access statistics we also must examine differences among Internet users. People encounter numerous hurdles during their online information-seeking behavior. In this paper, I focus on the likelihood that Internet users will make spelling or typographical mistakes during their online activities. Information seeking on the Web often requires users to type text into forms. Users sometimes make mistakes, which can hinder their browsing efficiency because they may get detoured to irrelevant sources or encounter errors. I draw on data collected from in-person observations with a diverse sample of 100 Internet users to see what explains their tendency to make spelling and typographical mistakes and the frequency with which they make such errors. I find that education level is a significant predictor of one's likelihood to make mistakes, suggesting that existing social inequalities translate into differences in online behavior.

Keywords: digital divide, digital inequality, errors, information seeking, online behavior, spelling, skill, Web use

Introduction

In the mid-1990s, concerns about a "digital divide" gained increasing attention with the release of the Falling Through the Net series by the National Telecommunications and Information Administration (1995; 1998; 1999; 2000). Data about the computer and Internet usages of a nationally representative sample of the population showed that the spread of information technology across different demographic groups was far from equal. The report showed that people in rural areas and those with lower levels of

[1] Robert Kauffman was the accepting senior editor.

education and income were less likely to be connected. Much attention has since been paid to this topic in an attempt to discern the specific contours of the digital divide along various dimensions of the socio-economic scale (DiMaggio et al. 2004; Loges and Jung 2001; Ono and Zavodny 2003; van Dijk 2005; Wilhelm 2000).

As the research in this area developed over the years, an increasing amount of work began focusing on refined approaches to the "digital divide" (Hargittai 2002b; Hargittai 2003a; Mossberger, Tolbert, and Stansbury 2003; van Dijk 1999; van Dijk 2005). These researchers argued that merely looking at differences in connectivity levels misses an important part of the puzzle. Refined measures of connectivity and use were necessary to understand the real contours of inequality with respect to Internet use. In addition to examining access differences, researchers started looking at the differences in the types of uses to which people put the medium (Howard, Rainie, and Jones 2001) and users' skills to make the most of their Internet usage (Hargittai 2002b; Hargittai 2003a; Mossberger, Tolbert, and Stansbury 2003).

An important component of online skills is the ability to search for information effectively and efficiently (Hargittai 2002b; Hargittai 2003a). A growing literature in the field of information science has looked at the details of people's online information-seeking behavior (e.g. Rieh 2004). Some of this work has focused on how people use directories and menu structures (Chen, Magoulas, and Dimakopoulos 2005; Yu and Roh 2002) while other research has focused specifically on the use of search engines (Cothey 2002; Spink et al. 2002; Spink, Wolfram, and Jansen 2001). Most of this work tends to explore how specific features of the services and systems used by participants influence their actions. This line of research has been less interested in identifying how the background of the users may relate to their behavior. Moreover, little systematic work has concentrated on the hurdles to efficient online information-seeking behavior. In this paper, I focus on one type of obstacle, namely, spelling and typographical mistakes. I do so by examining the actions of a diverse group of users in order to explore how different socio-economic variables may contribute to differences in users' online behavior.

Refining Approaches to The Digital Divide

Although the digital divide was at first a rare topic among academics, a literature review in social science databases shows that the number of research articles published about the digital divide has increased considerably over time, with more than 100 such papers published in each of the last few years. At first, most of the focus concentrated on explaining differences in connectivity statistics by population segments (Bimber 2000; Bucy 2000; Hargittai 2003b; Loges and Jung 2001; Norris 2001; Wilhelm 2000). The "divide" was understood as a simple binary measure of access versus no access, or at best use versus no use.

Moving the agenda forward, recent work has increasingly broadened the research program to focus on refined measures of access and use, including quality of access, context and intensity of use, types of utilization and user abilities (Attewell 2001; Bonfadelli 2002; DiMaggio et al. 2004; Hargittai 2002b; Hargittai 2004b; Howard, Rainie, and Jones 2001; Katz and Rice 2002; Mossberger, Tolbert, and Stansbury 2003; van Dijk 2005; Warschauer 2003; Wellman et al. 2002). In these investigations, the differences are no longer considered as a dichotomous property; rather, they exist on a

spectrum. In fact, DiMaggio et al. (2004) advocate the use of the term "digital inequality" instead of "digital divide" to reflect more accurately the varying levels of use. In this paper I continue the tradition of exploring refined measures of digital inequality. In particular, I focus on ways in which users' particular online actions may hinder the extent to which they can optimally benefit from their use of the medium.

Refined data about average users' online behavior show that while some activities are nearly universal (e.g. 93% of users say they have sent or received email), many activities are less widespread across all users (Madden 2003). Seventy-one percent have consulted news (Madden 2003) and 80% have looked for some type of health-related information online (Fox 2005). These activities are not equally distributed across the Internet-user population, however. For example, while 87% of those with a broadband connection at home sought some health information on the Web according to a survey conducted in 2004, only 72% of those with a home dial-up connection did so (Fox 2005). Also, Internet veterans (in the case of Fox's study, referred to as people who've been online for six or more years) are considerably more likely to have engaged in such activity (86%) than those who have only two to three years of online experience (66%). These figures suggest that certain attributes of users' Internet-related experiences (i.e. quality of connection, history of Web use) influence the types of activities people pursue online.

Of course, we would need more information about users to draw any conclusions regarding the independent effect of certain factors on people's online activities. Those who started using the Internet later and who don't have high-speed connections at home (e.g. lower income, lower education) may differ from others in various ways, which may then be related to their propensity to search for health information in the first place. Nonetheless, these relationships are important to note and suggest that growth in basic user statistics does not necessarily mean that everybody is taking advantage of the medium in similar ways. Precisely because those who have become users in the past year or two do not share the same demographics of early adopters, uses by veteran status may differ.

While people go online for a myriad of activities-from sending and receiving email to instant messaging-information retrieval constitutes a significant part of people's daily Web use. In particular, we know that the majority of users turn to search engines for at least some of their online information needs, and this particular online activity ranks number two in popularity behind email use for the average user (Fox 2002). The high rank of search engine use among online activities underscores the importance of the Internet as a source of information for users.

As the amount of information online has grown exponentially over the years, the need for tools to sift through the material has gotten larger. Search engines and portal sites have evolved to meet these needs (Hargittai 2004c), but they still require a certain level of understanding and skill to be used effectively (Hargittai 2002b). Although, in theory, the Internet may offer information on every imaginable topic, it is easy to get lost in the vastness of resources. If those in need of certain types of material are unable to find it, the mere availability of the content will not aide them. Thus, people's ability to find desired types of information is an important part of the medium's potential to contribute to their everyday needs and well-being, and ultimately improve their life chances.

A nuanced approach to digital inequality takes a critical look at how people are able to benefit from technologies once they have gained access to them. A look at people's ability to navigate online content will help to discern the extent to which users are able to benefit from use of the medium. I focus specifically on people's likelihood of making typographical and spelling mistakes during their online activities to gain a better understanding of whether this hurdle is randomly distributed among users or whether it has particular predictors.

Making Mistakes While Using the Web

Although making mistakes in typing while navigating online content may seem like a mere nuisance, evidence suggests that such errors can have direct financial consequences (Schemo 2004). This study reported on people finding bargains on the online auction site eBay by searching specifically for misspelled merchandise. Some users of the online service found that they could benefit from low-trafficked bids by looking up listings of items using common misspellings of certain words. Most people tend to search for products using correct spelling, so few bidders come across items that are listed under misspelled words. Therefore, those who find them can save substantial amounts by not having to bid against all possible interested parties. While some reap the benefits of others' mistakes, those making the errors lose out on the potential financial rewards of selling their merchandise.

Although some research has looked at the frequency of errors during users' information-seeking behavior, this work has focused exclusively on analysis of large log files (Cooper 2001; Kurth 1993; Peters 1993). While this is a fruitful approach to get an idea of how likely users are, in general, to make mistakes, this work has not gone the extra step to explore the predictors of making typographical and spelling mistakes. Since engaging in such behavior has consequences for how people are able to benefit from use of the medium, it is of interest to explore what factors may underlie such actions. In this paper, I take a detailed look at what predicts error-prone typing by Internet users.

Undoubtedly, people make more than spelling and typographical errors while performing actions online. They may click on links by mistake, they may fill out the wrong forms for their purposes, or they may type in the wrong URL when trying to access a specific Web site, among other possible errors. Here, I focus on a particular set of mistakes related to typing, because these are common occurrences, as data presented below will show. Moreover, these mistakes – especially the tendency to misspell words – are independent of Internet use per se. They are skills that are part of educational curricula from the start (i.e. there is considerable focus on spelling abilities from early grades in schools) and are independent of changing technologies and platforms (unlike the specifics of Web use that relate to links, forms or URLs). Yet they are relevant to making the most of the Web, as exemplified by the online auction case described above. Future research could explore other types of mistakes and see whether users who are prone to making one type are also more likely to run into other difficulties.

In the next section, I describe the data on which I base the subsequent analyses. I present the sampling and recruitment technique, descriptive statistics about the group of participants, how I administered the studies, and the coding scheme I used to aggregate the data. I follow by discussing the methods of analysis and the findings. After presenting the results, I discuss their implications for a refined understanding of the "digital divide" or digital inequality.

Data and Methodology

The Sample

I draw on data collected through in-person observations and interviews with 100 randomly sampled adult Internet users from Mercer County, New Jersey between the summers of 2001 and 2002.[2] Internet users were defined as people who go online for browsing the Web at least once a month and had done so in the past month. Respondents were asked to come to a university location for participation and were offered $40 for their time and effort (with additional compensation offered for babysitting or transportation costs). The response rate was 54%, considerably high given the type of active participation required of respondents.

Unlike other studies (McDonald and Spencer 2000; Wang, Hawk, and Tenopir 2000), the participants in this project represent a random sample of county residents and are therefore a diverse group of Internet users. They range in age from 18-81, about half (51%) are women (see Table 1 for details).[3] Participants' occupations range from real-estate agents, environmental policy analysts, blue-collar workers to office assistants, teachers, service employees and medical professionals in addition to students, unemployed and retired persons.

Table 1. Descriptive statistics about sample participants

	Mean	St. dev.	Median	Minimum	Maximum
Age	42.96	15.86	42	18	81
Education (a)	16.21	2.72	College	Less than high school	Ph.D.
Family income (b)	$98,394	$57,452	$80,000-89,000	$17,500-19,000	>$250,000
Number of years since first use of the Internet	6.28	3.38	6	0	16
Number of hours browsing the Web weekly	8.62	9.39	7	8 minutes	70 hours

Notes: (a) Education was collected as a categorical variable, which was then converted into years for the analyses. The mean and standard deviation are derived from the converted categorical values. (b) Household income was collected as a categorical variable, which was then converted into dollar values using the mid-point of the categories. I used $274,999.5 for the top category of "$250,000 or more".

The sample is representative of the average Internet user in the local county population regarding socio-economic factors (see Hargittai 2003a for a detailed discussion of the comparison). However, the county as a whole is one of the most well-to-do counties in the nation, so is not representative of the United States. Regarding the focus of this

[2] I purchased a random sample of residential addresses and sent letters to households to request participation. Soon after, follow-up phone calls were made to recruit a random adult member of the household for participation.

[3] 14% of respondents were minorities; seven African American, four Asian American, and three Hispanic people took part in the study. These numbers are too small to draw inferences about the effects of race and ethnicity on skill so such variables are not included in the analyses.

study, this suggests that findings concerning the relationship of education and income to searching abilities will likely be conservative.

The group is also diverse regarding Web-use frequency and history. Participants' Web use ranges from just a few minutes a week to over 30 hours weekly. The group is similarly diverse in its overall experience with the medium. One person went online the year of the study and an additional 13% had only used it for two years or less. However, many – 39%– of the subjects had been users for five to seven years. There are also several long-term users among the respondents, with 15% having had their first exposure to the Internet over a decade before the study was administered.

Data Collection

At the study sessions, the researcher first administered an oral questionnaire to collect background information about subjects' usual Web-use experiences and to establish a rapport with respondents.[4] Next, participants were asked to sit at a computer and perform a variety of tasks online by looking for various types of content. Subjects were given the choice of using a PC or a Mac, both of which were loaded with the three most popular browsing software applications (Internet Explorer, America Online, and Netscape Communicator) to allow respondents to replicate their usual online experience.[5]

The computers connected to the Internet on a high-speed university network line. A program was used to erase the browser and URL history on the computer so that each respondent started out with a clean slate and was not influenced by previous users' actions. The search sessions were recorded with a screen capture program that generated audio-visual files of the entire search session. After the observation session, an online survey collected data on subjects' demographic background. Hargittai (2002a) describes the methodology in more detail.

Respondents were asked to find different types of content reflecting various online activities that may be relevant to their everyday lives, but some material was included with which they would likely have little experience. The seventeen tasks included looking for information about political candidates, a health-related Web site, tax forms, product information, and some art and cultural content. The focus was purposefully not on trivia questions and there were myriad ways in which people could solve the tasks.

Coding

I coded the audio-visual recordings of the search sessions for all online actions. I created a coding and classification scheme that offers an exhaustive list of ways in which one may arrive at a Web page (Hargittai 2004a). This coding scheme allows the data to be tabulated in a way that facilitates subsequent aggregation of information. See Figure 1 for an example of coded actions. Whenever users made a typographical or spelling mistake while typing in a search query or filling out a form, I transcribed the

[4] I conducted 80 of the interviews; two female research assistants administered the remaining 20.
[5] No default page was set on browsers in order not to influence respondents' initial actions once online. The sessions were started off by the researcher asking the respondent to recall – if possible – the default homepage on the computer she uses the most.

exact wording into the coding sheet with the addition of "(sic)" to indicate that the mistake was in the original file and was not an error introduced during coding. A simple search for "sic" in the database of all coded actions provides a list of all instances in which users made mistakes. I then aggregated these for each user.

id	access	URL	time	task	scroll	search_term	link_text	error	from_h	form_info	select	button	auto-c	success
1097	61	www.healthscout.com	15.22	7	0		Health							
1097	21	www.healthscout.com	15.65	7	0	latoce (sic)						Health Scout Search		
1097	21	www.healthscout.com	16.02	7	0	lactoce intolerance (sic)								
1097	44	www.healthscout.com	16.15	7	2		Lactose Intolerance							1
1097	11	www.healthscout.com	16.77	8	2		Diet, Fitness, Looks							
1097	71	www.healthscout.com	17.32	8	0									
1097	44	www.healthscout.com	17.63	8	0		Food Allergy or Food Intolerance?							
1097	70	www.healthscout.com	18.48	8	0									

Figure 1. Example of coded actions

I decided whether an error was a typographical or a spelling mistake by looking at the specific instances. For example, the phrase "lost waller" instead of "lost wallet" was considered a typographical error, because the "r" is located right next to the "t" on the keyboard and "waller," does not sound at all like "wallet" so the respondent could not have meant to type that. In contrast, typing "lactoce" instead of "lactose" was considered a spelling mistake. Not only are the keys "c" and "s" not located next to each other on the keyboard, using a "c" in that word was likely intentional, as reading the word with that spelling results in a similar sounding term as the correct spelling. If a search term contained both a typographical and a spelling mistake then I coded it as having had both. An additional researcher coded the entire data set in order to establish inter-coder reliability. Table 2 reports the kappa values for the variables. As these are all very high (one is 83%, the others are all above 90%), the coding is deemed reliable and consistent for analysis.

Table 2. Inter-Coder Reliability Scores for Classifying Types of Mistakes

Variable	Agreement	Kappa	Standard Error
Made any spelling mistakes (binary)	98%	0.96	0.10
Number of spelling mistakes	96%	0.94	0.05
Made any typographical errors (binary)	96%	0.91	0.10
Number of typographical errors	91%	0.83	0.07
Made any mistakes (binary)	100%	1.00	0.10
Number of all mistakes	95%	0.94	0.05

The data set includes only the errors that respondents actually submitted to search engines or browsers in the location field. There is no record of mistakes that respondents corrected before clicking on the search button or pressing the <Enter> key, so the number of mistakes included here is a conservative representation of how frequently users engage in inaccurate typing and spelling practices. Table 3 presents examples from the dataset of both spelling mistakes and typographical errors committed by users in the study and submitted to search engines or in the location bars of browser windows, a method in which many users engaged interchangeably with the use of search engines.

Table 3. Examples of Actual Typographical and Spelling Mistakes Made by Respondents in the Study

Examples of typographical errors	Examples of spelling mistakes
1040 tak form	lactose intolerant recipy
stoloen property	new jersey govenor election
presidentrai elections	Lactose Intolerence
Bush/Gore Apbortion	christian science moniter
princeeeton packet	volenteer opertunities

Methods of Analysis

I use two types of dependent variables in the analyses. First, I look at the likelihood of making mistakes at all. I run a logistic regression on the binary variable that indicates whether the respondent had made any typographical or any spelling mistakes. I also create a summary variable to look at the likelihood of making any type of mistake. Because I have nuanced information regarding the frequency with which respondents committed such errors, I also perform OLS regression analyses on the total number of mistakes made by respondents separately for typographical and spelling errors, as well as for a summary variable. Table 4 presents the correlation coefficients for the variables included in the model. There are no prohibitively high correlations, suggesting that there are no concerns of multicollinearity.

I use information about participants' demographic characteristics and their experience with the medium to explain the outcome variables. I check the relationship of age, gender, education and income with the dependent variables. I also look at the importance of using a computer at home, and respondents' experience using the Internet. I measure Internet use experience with two variables: a) the amount of time spent on the Web weekly and b) the number of years having used the Internet.

Table 4. Correlation coefficients of variables in the model

	Female	Age	Education	Income	Web time	Use years
Age	-0.0359 (0.7232)					
Education	-0.0571 (0.5728)	0.2952 (0.0029)				
Income	-0.0250 (0.8048)	0.0321 (0.7510)	0.0728 (0.4718)			
Web time	-0.1401 (0.1644)	-0.2221 (0.0264)	-0.0099 (0.9222)	-0.0258 (0.7989)		
Use years	-0.0044 (0.9656)	-0.2933 (0.0031)	0.1166 (0.2481)	0.0633 (0.5313)	0.3799 (0.0001)	
Freedom to use Net at work	-0.1032 0.3071	-0.3544 0.0003	0.1136 0.2605	0.0971 0.3364	0.0914 0.3656	0.3800 0.0001

Findings

Users make spelling mistakes and typographical errors both during searches and while typing in the addresses of Web sites. This is a very common hindrance to efficient online navigation. Almost a quarter (23%) of respondents made at least one typo during their search session, over half (52%) made at least one spelling mistake, and almost two-thirds (63%) did one or the other. Among those who made mistakes, 35% only made one mistake, but over a quarter of those making mistakes (17 participants) made four or more during their entire session. Table 5 gives details about the frequency of the two types of mistakes by respondents in the sample.

Table 5. Frequency of Mistakes by Respondents in the Sample

	Percentage of respondents making mistakes		
Number of Mistakes	Spelling	Typographical	Both types
0	48	65	35
1	20	20	19
2	10	10	12
3	10	1	11
4	6	3	13
5	2	1	5
6	2	0	3
7 or more	2	0	2

Next, I turn to exploring users' propensity to make mistakes during their online activities. In Table 6, I present results of a logistic regression explaining who is more likely to make spelling and typographical mistakes during their online actions. I use logistic regression to estimate the likelihood of a user making any typographical errors, any spelling mistakes and a summary of whether they make either type of mistakes.

Table 6. Logistical Regression Predicting Characteristics of Users Who Make Typographical Errors, Spelling Mistakes and Either Type of Mistakes During Their Search Sessions

	Made Typographical Error	Made Spelling Mistake	Made Either a Typo or a Spelling Mistake
Female	0.135 (0.466)	-0.325 (0.439)	-0.209 (0.473)
Age	0.025 (0.018)	0.021 (0.016)	0.029 (0.019)
Education	-0.066 (0.091)	-0.195* (0.091)	-0.195* (0.099)
Family Income	0.864* (0.404)	-0.379 (0.346)	0.251 (0.371)
Freedom to use Net at work	-1.172* (0.538)	-0.366 (0.490)	-1.153* (0.542)
Time on Web/week	-0.301 (0.300)	0.158 (0.278)	-0.059 (0.303)
User years	0.745 (0.523)	-0.088 (0.488)	0.028 (0.548)
Intercept	-10.773* (4.733)	6.887 (4.246)	0.636 (4.517)
N	100	100	100
Wald χ^2	15.131	10.501	16.261
Note: *** p<0.001; ** p<0.01; * p<0.05; # p<0.1; two-tailed test			

Results suggest that one's level of income is related to the tendency to make typographical mistakes, while education is a significant predictor of spelling mistakes. Curiously, the higher one's income the more likely one is to make typos. Regarding education, the relationship is in the expected direction. Those with more years of schooling are less likely to make spelling mistakes than those who have fewer years of education. We also find that having the freedom to use the Web on the job works to one's advantage when it comes to making typographical errors, as those with more freedom are less likely to make such mistakes. However, the experience variables do not influence the likelihood of making spelling mistakes. In these analyses we find no relationship between making mistakes and gender or age.

Next, I look at more refined data regarding the number of mistakes people made during their online sessions. In Table 7, I present the results of OLS regression models on the frequency of mistakes made. Similarly to the previous analyses, I break down the mistakes by type, first looking at typographical errors and spelling mistakes separately and then including the sum of the two in one model.

Table 7. OLS Regression Predicting Number of Times Users Make Typographical Errors, Spelling Mistakes and All Mistakes Combined During Their Online Sessions

	Number of Typographical Errors	Number of Spelling Mistakes	Number of Typos and Spelling Mistakes
Female	-0.003 (0.207)	-0.305 (0.354)	-0.307 (0.423)
Age	0.006 (0.008)	0.023# (0.013)	0.030# (0.016)
Education	-0.041 (0.041)	-0.192** (0.070)	-0.233** (0.084)
Family Income	0.301# (0.164)	-0.486# (0.280)	-0.185 (0.334)
Freedom to Use Net at work	-0.366 (0.237)	-0.313 (0.406)	-0.679 (0.485)
Time on Web/week	-0.100 (0.131)	0.210 (0.223)	0.109 (0.267)
User years	-0.080 (0.229)	-0.357 (0.392)	-0.437 (0.468)
Intercept	-1.860 (1.946)	9.526** (3.325)	7.667# (3.970)
N	100	100	100
R2	.113	.183	.192
Adjusted R2	.045	.120	.130
Note: *** p<0.001; ** p<0.01; * p<0.05; # p<0.1; two-tailed test			

Results suggest that income displays a similar relationship with frequency of typos as it did with the tendency to make typographical mistakes. However, freedom to use the Web at work is not significant in this model. Regarding number of spelling mistakes, we now see a significant relationship with age. Older users make more spelling mistakes

than younger users. Also, as with the tendency to make spelling mistakes, education is also a negative predictor of how often people make such mistakes. We also find that those with a higher income are less likely to misspell words during their online navigation.

Discussion

Overall, we find no gender differences in Web users' tendency to make typographical or spelling mistakes, and the effects of age are constrained to the frequency of spelling mistakes. Rather, we find some significant relationships with socio-economic factors. As one might expect, those with more years of schooling make fewer mistakes. This finding has implications for discussions about digital inequality, as it suggests that those already in more privileged positions are able to navigate the Web more seamlessly than those who have fewer years of education. Amount of experience with the Internet – both regarding number of years online and amount of time spent on the Web – does not seem to make a difference in the tendency to commit and frequency of mistakes. There is one exception: freedom to use the Web at work leads to a lower likelihood of making typographical mistakes. This finding suggests that so-called autonomy of use – freedom to use the medium when and where one wants to – can have positive repercussions for one's overall experiences with the medium.

Although several search engines have tried to respond to users' typographical and spelling mistakes by developing cues for users about errors submitted in search queries, these do not always help. Few users tend to notice the alternative – i.e. correct – spelling offered by search engines on the top of their results lists. The search engine Google, for example, signals a possible spelling mistake or typographical error by placing the words "Did you mean" on top of the results list and suggesting a correct version of the word or phrase. In order to take advantage of the proposed correct spelling, the user need only click on this text, which is a live link that leads to the search results of the correctly spelled query. However, users tend to skip over this despite its prominence. In this study, among the 37 users who searched with Google at some point during their session and also made mistakes, none clicked on this link at any point during their session.[6]

Part of the reason for not taking advantage of this option may be that often a search engine will return results even when something in the query is misspelled. This happens when Web pages mirror the same spelling mistakes made by the user. Although these pages can be relevant to the intended query, they are not necessarily optimal sources if they contain typographical or spelling errors themselves. Figure 2 illustrates some of these cases.[7]

Search engines have tried to introduce various measures to deal with users' typographical and spelling mistakes. Sometimes search engines automatically redirect

---

[6] The coding scheme used to generate information about people's actions on the Web sites includes a code for having clicked on such a "Did you mean?" link. However, this code never shows up in the data set suggesting that no one in this study ever clicked on the link. [Hargittai, Eszter. 2004a. Classifying and Coding Online Actions. *Social Science Computer Review* 22 (2).

[7] All screen shots presented here are derived from actual study sessions.

to what the underlying technology assumes the user meant by his original query. Although this may be helpful in some cases, it may be too much intervention in others. If the user entered a purposefully less traditional spelling, then it will be more difficult to get the results to that query. Few users know the intricate algorithms that search engines use, so it would be hard for them to know how to force a search engine to take their intended spelling in unorthodox cases.[8]

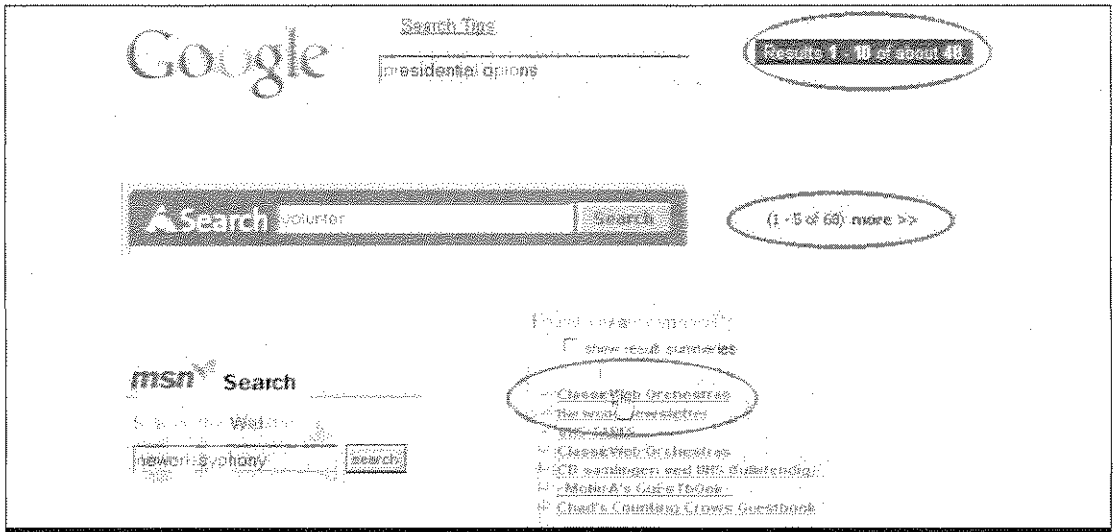

Figure 2. Examples of Typographical and Spelling Mistakes during the Use of Google, AOL Search and MSN Search All of Which Led to a List of Results Despite the Errors

Conclusion

The data presented in this paper – based on the online information-seeking behavior of a diverse sample of Web users – suggest that typographic and spelling mistakes are common during people's browsing behavior. Analyses of the data imply that the propensity to make such errors is not randomly distributed among users. Rather, users' education level exhibits a statistically significant relationship with a tendency to make spelling mistakes and the frequency with which users make such mistakes. Those who have more years of schooling are less likely to run into the difficulty that results from misspelling terms. This finding suggests that online behavior is not independent of existing social inequalities. The Internet has the potential to level the playing field among people from differing socio-economic backgrounds by making vat amount of information available around the clock. However, if those in less privileged positions lack some of the know-how to use the medium efficiently, they may be less likely to benefit from the Web than those with better skills to navigate online content.

[8] For example, by adding a + sign right in front of a term forces the search engine Google to include it in the search as is. This is the one way one can make sure that general terms such as "the" and "and" are included in queries. Otherwise, Google disregards such generic terms when running a search. However, data collected during this study suggest that very few users turn to such refined strategies when using search engines.

In the introductory section and the review of the literature, I noted the increasing emphasis on research that examines differences among Internet users and not simply differences between users and non-users. The findings presented in this paper underscore that even among users, inequalities remain. The traditional digital divide approach assumes that once people gain access to digital technologies, all inequalities will be addressed. The above findings challenge that simplistic approach to digital inequality. Simply having access and being a user does not mean that all hurdles to efficient uses of the medium have been removed.

Internet use does not happen in isolation of other skills on which people rely in their everyday lives. Despite technological solutions to hurdles such as people's tendency to make spelling mistakes, many users' experiences are nonetheless influenced by such errors. In order to increase the Internet's potential to level the playing field among those with divergent backgrounds regarding spelling and computer-use skills, systems designers need to develop and introduce even better tools and services that help users – users from diverse backgrounds – sidestep these hurdles.

Acknowledgements

I thank the guest editors and the anonymous reviewers for their very helpful comments. I am grateful to Amanda Hinnant for her help with coding. I would like to acknowledge generous support from the Markle Foundation and NSF grant #IIS0086143. The project has also been supported in part by a grant from the Russell Sage Foundation, and through a grant from the Pew Charitable Trusts to the Center for Arts and Cultural Policy Studies, Princeton University. I am also grateful to the Dan David Foundation for its support.

References

Attewell, P. 2001. The first and second digital divides. Sociology of education 74 (3):252-259.

Bimber, B. 2000. The Gender Gap on the Internet. Social Science Quarterly 81 (3):868-876.

Bonfadelli, Heinz. 2002. The Internet and Knowledge Gaps. A Theoretical and Empirical Investigation. European Journal of Communication 17 (1):65-84.

Bucy, E. 2000. Social Access to the Internet. Harvard International Journal of Press/Politics 5 (1):50-61.

Chen, Sherry Y., George D. Magoulas, and Dionisios Dimakopoulos. 2005. A Flexible Research Interface Design for Web Directories to Accomodate Different Cognitive Styles. Journals of the American Society for Information Science and Technology 56 (1):70-83.

Cooper, Michael D. 2001. Usage Patterns of a Web-Based Library Catalog. Journal of the American Society for Information Science and Technology 52 (2):137-148.

Cothey, V. 2002. A Longitudinal Study of World Wide Web Users' Information-Searching Behavior. Journal of the American Society for Information Science and Technology 53 (2):67-78.

DiMaggio, Paul, Eszter Hargittai, Coral Celeste, and Steven Shafer. 2004. Digital Inequality: From Unequal Access to Differentiated Use. In Social Inequality, edited by K. Neckerman. New York: Russell Sage Foundation.

Fox, Susannah. 2002. Search Engines:
A Pew Internet Project Data Memo. Washington, D.C.: Pew Internet and American Life Project.
Fox, Susannah. 2005. Health Information Online. Washington, D.C.: Pew Internet and American Life Project.
Hargittai, Eszter. 2002a. Beyond logs and surveys: In-depth measures of people's Web use skills. Journal of the American society for information science and technology perspectives 53 (14):1239-1244.
Hargittai, Eszter. 2002b. Second-Level Digital Divide: Differences in People's Online Skills. First Monday 7 (4).
Hargittai, Eszter. 2003a. How Wide a Web? Inequalities in Accessing Information Online. Ph.D., Sociology Department, Princeton University, Princeton, NJ.
Hargittai, Eszter. 2003b. The Digital Divide and What To Do About It. In The New Economy Handbook, edited by D. C. Jones. San Diego, Calif: Academic Press.
Hargittai, Eszter. 2004a. Classifying and Coding Online Actions. Social Science Computer Review 22 (2).
Hargittai, Eszter. 2004b. Internet Access and Use in Context. New Media & Society 6 (1).
Hargittai, Eszter. 2004c. The Changing Online Landscape: From Free-for-All to Commercial Gatekeeping. In Community Practice in the Network Society: Local Action/Global Interaction, edited by P. a. D. S. Day: Routledge.
Howard, Philip N., Lee Rainie, and Steve Jones. 2001. Days and Nights on the Internet: The Impact of a Diffusing Technology. American Behavioral Scientist 45 (3):383-404.
Katz, J.E., and R.E. Rice. 2002. Social Consequences of Internet Use: Access, Involvement and Interaction. Cambridge, MA: MIT Press.
Kurth, M. 1993. The Limits and Limitations of Transaction Log Analysis. Library Hi Tech 42:98-104.
Loges, William E., and Joo-Young Jung. 2001. Exploring the Digital Divide: Internet Connectedness and Age. Communications Research 28 (4):536-562.
Madden, Marry. 2003. America's Online Pursuits. Washington, D.C.: Pew Internet and American Life Project.
McDonald, Sharon, and Linda Spencer. 2000. Gender Differences in Web Navigation: Strategies, Efficiency, and Confidence. In Women, Work, and Computerization: Charting a Course to the Future, edited by E. Balka and R. K. Smith. Boston: Kluwer Academic Publishers.
Mossberger, Karen, Caroline J. Tolbert, and Mary Stansbury. 2003. Virtual Inequality: Beyond the Digital Divide. Washington, D.C.: Georgetown University Press.
National Telecommunications and Information Administration. 1995. Falling Through the Net: A Survey of the "Have Nots" in Rural and Urban America. Washington, DC: NTIA.
National Telecommunications and Information Administration. 1998. Falling Through the Net II: New Data on the Digital Divide. Washington, DC: NTIA.
National Telecommunications and Information Administration. 1999. Falling through the net: Defining the digital divide. Washington, DC: NTIA.
National Telecommunications and Information Administration. 2000. Falling Through the Net: Toward Digital Inclusion. Washington, DC: NTIA.
Norris, P. 2001. Digital Divide: Civic Engagement, Information Poverty and the Internet in Democratic Societies. New York: Cambridge University Press.
Ono, Hiroshi, and Madeline Zavodny. 2003. Gender and the Internet. Social Science Quarterly 84 (1).

Peters, T.A. 1993. The History and Development of Transaction Log Analysis. Library Hi Tech 42:41-66.

Rieh, Soo Young. 2004. On the Web at Home: Information Seeking and Web Searching in the Home Environment. Journals of the American Society for Information Science and Technology 55 (8):743-753.

Schemo, Diana Jean. 2004. In Online Auctions, Misspelling in Ads Often Spells Cash. The New York Times, January 28, 2004, 1.

Spink, A., B.J. Jansen, D. Wolfram, and T. Saracevic. 2002. From E-Sex to E-Commerce: Web Search Changes. IEEE Computer 35 (3):107-109.

Spink, Amanda, Dietmar Wolfram, and Major B. J. Jansen. 2001. Searching the Web: The Public and Their Queries. Journal of the American Society for Information Science and Technology 52 (3):226-234.

van Dijk, Jan A.G.M. 1999. The Network Society: Social Aspects of New Media. London: Sage.

van Dijk, Jan A.G.M. 2005. The Deepening Divide. London: Sage Publications.

Wang, Peiling, William B. Hawk, and Carol Tenopir. 2000. Users' Interaction with World Wide Web Resources: An Exploratory Study Using a Holistic Approach. Information Processing and Management 36:229-251.

Warschauer, M. 2003. Technology and Social Inclusion. Cambridge, Mass: MIT Press.

Wellman, B, A.Q. Haase, J. Witte, and K Hampton. 2002. Capitalizing on the Internet: Social Contact, Civic Engagement, and Sense of Community. In The Internet in Everyday Life, edited by B. Wellman and H. Haythornthwaite. Oxford: Blackwell.

Wilhelm, Anthony G. 2000. Democracy in the Digital Age: Challenges to Political Life in Cyberspace. New York: Routledge.

Yu, Byeong-Min, and Seak-Zoon Roh. 2002. The Effects of Menu Design on Information-Seeking Performance and User's Attitudes on the World Wide Web. Journal of the American Society for Information Science and Technology 53 (11):923-933.

About the Author

Eszter Hargittai is Assistant Professor in the Department of Communication Studies and Faculty Fellow at the Institute for Policy Research at Northwestern University. The focus of her research is on how people use the Web in their everyday lives and in particular, how differences in Internet use may contribute to social inequality. She received her Ph.D. in Sociology from Princeton University where she was a Woodrow Wilson Scholar. Information about her work is available at www.webuse.org.

Journal of the Association for Information Systems
ISSN: 1536-9323