INSEAD

The Business School for the World®

Faculty & Research Working Paper

Empirical Evidence for the Role of the Domain Name Itself in Website Performance

Karan GIROTRA
Karl ULRICH

2010/98/TOM

Empirical Evidence for the Role of the Domain Name Itself in Website Performance

Karan Girotra*

Karl Ulrich**

November 2010

* Assistant Professor of Technology and Operations management at INSEAD, Boulevard de Constance 77305 Fontainebleau Cedex Ph: 33 (0)1 60 72 91 19
Email: karan.girotra@insead.edu

** CIBC Professor, Professor of Operations and Information Management at The Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, Pennsylvania, 19104, USA Email: ulrich@wharton.upenn.edu

Empirical Evidence for the Role of the Domain Name Itself in Website Performance

This paper provides the first large-scale empirical evidence of the association between specific properties of internet domain names and website performance. We analyze over one million internet domain names, linking their phonological and morphological attributes to the realized demand for their associated websites. We test hypotheses related to how the names sound, how they look, their ease of recall, and the likelihood that they will be typed correctly. We find that certain attributes of names are associated meaningfully and significantly with the demand realized by a website. The websites with the highest demand have names that are short, include dictionary words, avoid punctuation symbols, and use numerals. The use of phonemes associated with disgust is negatively associated with performance for most websites, but positively associated with performance for adult sites. Some of these results from the on-line world are likely to hold off line, while some are not. These findings can be used in conjunction with other criteria as part of the selection process for names.

Key words: domain names, names, brands, naming, branding, euphony, phonology, dot-com, linguistics
History: This is the first version of this paper – November 2010

---

1 Introduction

The name of a product, service, or organization is the label used to identify the collection of elements comprising that entity. A rich body of research in psychology, marketing, and linguistics convincingly shows that names are important for the way humans make sense of the world and the ways in which they store and retrieve information (e.g., (Keller & Lehmann, 2006), (Macnamara, 1984), (Margolis & Laurence, 1999)). In the domain of business, hundreds of scholars have studied brands and branding, linking names of products, services, and companies to consumer judgments and the financial value of firms ((Aaker, 1991), (Keller, 2003)).

Missing from this research has been large-scale empirical evidence for the value of attributes of the name itself. What characteristics of a name are associated with higher levels of business performance? The most significant barrier to answering this question empirically is obtaining a large enough data set with a reasonable proxy for the performance of the entity. Domain names, used as labels for websites, overcome this barrier, because rankings of site traffic are available for millions of websites. In this study, we exploit unique properties of internet domain names to test the relationship between the morphological and phonological properties of the names themselves and the demand realized by the websites identified by those names. Some of these relationships are specific to domain names, but we believe some may be more general, and are likely to hold in the off-line world.

Domain names (e.g., *ebay.com*) are labels defining the scope of administrative authority on the internet. The top-level domain names on the internet are the right-most portion of a name and include, for example, .com, .net, .edu, .org, and .gov. The second-level domain within a domain name may generally be registered by an individual or organization and often corresponds to a brand, company, or organization name. For example, the domain name ebay.com is comprised of the top-level domain ".com" and the second-level domain "eBay," which is registered to Ebay, Inc.

Because domain names are essentially addresses for the internet sites of their registrants, they must be unique. Valid secondary domains are strings of up to 63 ASCII characters drawn from a set of 37: the 26 letters of the English alphabet, the numerals 0 through 9, and the hyphen "-". In western economies, essentially the same set of characters is used to construct names in the off-line world, although some names include additional special characters and spaces. For example, the off-line name "Marks & Spencer plc." corresponds to "marksandspencer.com" on the internet.

Domain names have emerged at the beginning of 21st Century as quite important in society. The internet economy involves about $1.5 trillion of the global economy and engages more than one-fourth of the population of the planet ((Atkinson, Ezel, Andes, Castro & Bennett, 2010)). Domain names are the primary scheme by which this economy is organized. For an internet-based organization (e.g., YouTube), the domain is often central to the marketing of the organization. For instance, the domain name serves an internet retailer in much the same way as both the brand and the physical location do for a bricks-andmortar retailer. For a business with significant operations in the physical world, the internet presence is typically still critical for marketing, communications, and transactions.

An entire service industry is devoted to naming and branding ((Morris, 2004), (Frankel, 2005)). Most of the firms in this industry generate and evaluate domain names as well as traditional brand names. As with other business and marketing activities, some organizations perform branding and naming tasks in house and some contract with consultants.

The internet provides a unique opportunity to test empirically relationships between names and performance, a quest that has been largely elusive in the off-line world. This is partly because domain names are unique, avoiding the problems of disentangling *Kohl's* the retailer from *Kohl's* the steel supplier, but mostly because rankings of website traffic are available for over a million entities, which provides enough statistical power to even test for subtle effects.

The primary question addressed by this paper is whether phonological and morphological properties of a domain name are associated with the level of website traffic, and if so, how significant are the effects.

This paper is organized as follows. In Section 2 we review the related literature. Section 3 provides theory and develops our hypotheses. Section 4 describes the data. Section 5 reports on our analysis and hypothesis tests. Section 6 discusses the results and their implications. Section 7 provides concluding remarks, including limitations of this study.

2 Literature

We have already referred to the large body of literature on branding and naming. However, as far as we know, there is essentially no prior literature on domain names per se. As a result, we bring to bear prior research from disparate related questions. Much of that literature addresses the role of names in consumer judgments, and in related psychological phenomena such as recall.

In the field of marketing, Minton was one of first to document and discuss the meanings of trade names ((Minton, 1958)) embodying components such as –master, -matic, and -rama. This tradition continues today, with one example in (Sebba, 1986). The semantics of a name and/or of the words used in constructing a name are clearly significant in determining the inferences people make from the names, although the semantic processing of names is a complex phenomenon, which is only partially understood (e.g., (Hennessey, Bell & Kwortnik, 2005)).

A person's response to a name is at least partially a *thin-slice judgment* – an immediate assessment based on very limited information. (Peracchio & Luna, 2006) provide a review of the role of thin-slice judgments in consumer behavior more generally. The main idea is that when a user must make a decision based on limited information, then any features of the alternatives under consideration, even the most trivial, are likely to factor into that decision. In other words, first impressions matter. For instance, thin-slice judgments on the web that arise from cues such as the characteristics of in-bound and out-bound hyperlinks have been shown to influence customer perceptions of trustworthiness (Stewart, 2006), thin-slice judgments about the appearance of web pages have been shown to be important in the way users assess the security, ease-of-use, and trustworthiness of sites ((Chiravuri & Peracchio, 2003); (Haried & Zahedi, 2006)). Because exposure to a domain name is often a key part of the initial stimulus leading a user to a website, thin-slice judgments of names are likely to be important.

Memory and recall play a key role in the relationship between domain names and website performance. If people can not remember a name after exposure, then they are less likely to successfully visit the associated website. (Tehan & Tolan, 2007) synthesize prior research on the relationship between word length and recall, and conduct experiments to define the relationship between word length and cognitive performance for short-term and long-term memory tasks, serial recall, and free recall. Names are particular types of words, and so some of these findings are likely to be relevant to the performance of domain names. (Lowrey, Shrum & Dubitsky, 2003) found that recall for brand names increases for *unusual spelling*, *semantic appositeness* (i.e., fit between brand and attributes of the entity), *initial plosives* (i.e., hard initial consonants), *paronomasia* (i.e., puns and word plays), and decreases for *blending* (i.e., composing names from two words).

A significant body of literature links phonological properties of words, including names, with human judgments about the objects the names refer to. (Alter & Oppenheimer, 2006) show in an experimental setting that subjects expect companies whose names are easier to pronounce to perform better financially. They also show empirically that first-day share price for initial public offerings is positively associated with the fluency of pronunciation of the three-letter ticker symbol for the stock. (Song & Schwarz, 2009) find that subjects assess the riskiness of food additives and amusement park rides as higher when their names are more difficult to pronounce.

There is an extensive literature in psychology and marketing on *phonetic symbolism*, the association between specific phonemes and attributes of the things the associated names refer to (e.g., (Lowrey, 2007), (Klink, 2000)). For example, (Yorkston, Menon & Mick, 2004) found that ice cream named *frosh* is likely to be perceived as smoother than ice cream named *frish*.

There is little if any academic research on the *usability* of domain names, the ease with which they can be typed or spelled. There is some evidence, however, that mistyping is extremely common, as this phenomenon is widely exploited by phishing scams and other security attacks ((Edelman, 2008)). Usability, in this narrow sense, is likely to determine, in part, website performance.

3 A Theory of the Performance of Domain Names

The string that makes up a domain name can be mapped to morphemes, minimal units of meaning, (e.g., e-bay) or phonemes, minimal units of sound (e.g., ee-b-ay). Some morphemes are *unbound* and can be used alone (e.g., bay) and some are *bound* and must be connected to other morphemes to create meaning (e.g., e-). Some names are comprised of dictionary words (e.g., "Apple" and "Mattress Giant"), some are proper nouns (e.g., "Dell"), some are entirely synthetic (e.g., "Zocor"), and some are acronyms (e.g., "IBM"). These categories are not perfectly disjoint, and hybrids like *eBay* are common. (See (Rivkin & Sutherland, 2004) for a more nuanced typology of brand names)

In some cases, domain names may be phrases or sets of words, often intended as descriptive labels for human users or for automated search engines. For example, the domain name i-am-bored.com is a sentence with words separated by hyphens, and the associated website is a user-generated list of websites intended to mitigate boredom.

Users typically access the information on a website via a web browser, a software application that can send and retrieve information to and from another computer on the internet and that can display information encoded with one of the standard protocols (e.g., html). There are two pathways by which users can get to a website: (1) they can directly type the domain name (or a URL that includes a page address or file name along with the domain name) or (2) they can click on a hyperlink provided by a search engine or another website. Some search engines, notably Google, allow the user to skip the intermediate clicking step by electing to be taken directly from search terms typed in the search engine to the webpage of the first search result (e.g., the "I'm feeling lucky" button on Google).

For the direct-type pathway, the domain name itself figures prominently in the likelihood that a user exposed to that name will actually arrive at the corresponding website. If the name is incorrect, the user will generally not arrive at the desired location, although if the typed domain name does not exist at all, some browsers may redirect to a webpage that suggests likely alternatives for the incorrect domain. For the search pathway, the name itself is one of the factors that search engines use to construct a page ranking, and since the name itself usually appears in search results, it can influence the likelihood that a user will click on a particular hyperlink. Thus, even for the search pathway, the name itself is quite important in connecting a user to an intended destination on the internet.

The overall semantics of a domain name are obviously important. The name *jiffylube.com* communicates very specific messages about the service (lubrication) and the benefit proposition (fast—in a jiffy), and *wellsfargo.com* inherits meaning from a historical brand. Yet, beyond the linguistic meaning of the morphemes in the name, how might attributes of the domain relate to the likelihood that a user visits a website?

We borrow loosely from the general logic of the "awareness, availability, and preference" models from marketing to develop some theoretical relationships. For consumers to purchase a physical good, they must be aware of the product, it must be available to them, and they must prefer the good to the alternatives (Figure 1). The analogous elements for a person visiting a website are as follows. First, a user must be aware of the site. Second, the user must prefer the destination to the alternatives. Third, the user must be able to successfully navigate to reach the intended destination, avoiding coding defects and other obstacles. (In most of the world, websites are essentially "available" to everyone with an internet connection, so we substitute the concept of navigation yield for availability.) We treat these factors in turn.

3.1 Awareness

There are two basic ways a user can be aware of a website. First, they can recall it from exposure elsewhere or from prior use. If they have used the site previously, they may have bookmarked it to allow them to more easily return to the site, and therefore need only recall that they bookmarked the site and recognize the site from a list of bookmarks. Second, they can be referred to the site from some other webpage, either a referring content page or a search engine results page.

Shorter words have better recall when a list of words is being recalled, possibly because rehearsal of order is faster with shorter words. Longer words seem to be more easily recalled in isolation, but these results were found for recognition, not for unaided recall including spelling ((Caplan, Rochon & Waters, 1992), (Tehan & Tolan, 2007)). Taken together, the effect of the length of domain names on performance remains an open question.

In laboratory studies, brands with an initial stop plosive (i.e., the stop consonant phonemes P, K, T) have been shown to be better recalled ((Vanden Bergh, 1990), (Vanden Bergh, Collins, Schultz & Adler, 1984)).

Unusual spellings have been shown to enhance recall ((Lowrey, Shrum & Dubitsky, 2003)) after visual exposure. However, we would expect unusual spellings to create coding defects in recall after only auditory exposure. Internet domain names sometimes include numerals, which we believe constitutes an unusual spelling. This practice may enhance recall, although it may introduce coding defects as well.

Dictionary words are more easily recalled than nonsense words ((Laxon, Masterson & Coltheart, 1991), (Coltheart, Patterson & Leahy, 1994)). For example, remembering the string “babyfood” is considerably easier than remembering its anagram “dfyboaob,” presumably because strings comprised of dictionary words can be coded in memory as words and not as an arbitrary sequence of characters. A fictitious word, say “fabybood,” is comprised of fewer dictionary words (“by” and “boo”), yet is more “word like” than a random string of characters. Thus we would expect dictionary words to be easiest to recall, with word-like strings somewhat easier to recall than arbitrary sequences of letters. Few domain names are as random as “dfyboaob” but domain names do vary widely in their use of dictionary words.

In sum, we would expect recall, and therefore awareness, to be influenced by the length of the name, in both characters and syllables; the use of the initial plosives P, K, and T; the use of numerals in the name; and the fraction of the name comprised of dictionary words.

3.2 Preference

An interesting branch of research in linguistics and marketing has probed the association between phonological properties of words and the attributes of the things they describe. The basic finding has been that people infer attributes of things from the sounds of words used to describe those things. For example, across many cultures and languages, an object called a “tikiti” is more likely to be considered sharp and pointy than an object called a “swall.” (Klink, 2000,Klink, 2003) showed that particular consonants (stop vs. fricative) are associated with particular attributes (weight, speed, etc.) (Yorkston, Menon & Mick, 2004) found that sounds in a brand name can influence perceptions of the physical properties of the associated products.

In most of this research, the associations between phonological properties of names and the underlying properties of the object have been established for specific attributes (e.g., speed, smoothness, weight, caloric content).[1] These attributes are not uniformly positive or negative; their valence would depend on the intended use of the brand. Tikiti might be a nice name for knives, but a bad name for pillows. The exception to this observation about valence is sounds associated with *disgust*. The [uh] back vowel sounds (e.g., blunder, bungle, muck, and yuck) and [ew] (e.g., puke) sounds are associated with disgust ((Jespersen, 1922), (Jakobson & Waugh, 2002)). In a study of political elections, (Smith, 1998) found validation for a relationship between disgust sounds in candidates' names and losing elections. Since our study looks at names for a wide variety of types of entities, we can only test whether phonological properties are generally associated positively or negatively with performance.

In addition to invoking meaning in some cases, certain sounds appear to require less effort to pronounce and may therefore be preferred. We limit our conjectures to the starting phonemes in names, and hypothesize that the basic categories of phonemes will be preferred differently, and that this preference will be associated with differences in performance of the names. We classify initial phonemes into nine categories: Stop Plosive Consonants (SP), Nasal Consonants (NC), Fricative Consonants (FC), Lateral Consonants (LC), Glides (GL), Combination Consonants (CC), Simple Vowels (SV) and Diphthongs (DV). Phonemes included in each of the basic categories are provided in Table 8 in the Appendix.

In addition to specific phonemes, three more general attributes of how names sound may relate to preference and/or to recall: *alliteration*, *consonance*, and *assonance*. Alliteration is the repeating of a sound at the beginning of successive words (e.g., Crazy Cone). Consonance is the repetition of consonant sounds in a word (e.g., total). Assonance is the repetition of vowel sounds in a word (e.g., billet). We know of no comprehensive studies testing preference or recall for these three attributes in words.

[1]The associations between language and attributes of the things described goes beyond phonological properties. Song and Schwarz (2008) found that subjects associated difficulty in reading instructions with difficulty in completing the task described by those instructions. In a subsequent study (Song and Schwarz 2009) they found that subjects associated names that were difficult to pronounce with the riskiness of the activities associated with those names.

However, they are mentioned enough in the branding literature that we can hypothesize that they may have a general positive or negative effect on either or both of recall and preference.

In sum, we expect the disgust phonemes to negatively affect preference for most categories of entities. We have no a priori expectations about the direction of a relationship between other phonological properties, but we would expect the features of the starting phonemes to have some effect, and we would expect alliteration, consonance, and assonance to also have meaningful general effects.

3.3 Navigation Yield

Even if someone recalls the name of a website and wants to go there, he or she may not arrive at the site. We call the success fraction in reaching the website, *navigation yield*. One failure mode is coding defects in how the name is typed. Some of these defects are the result of miscoding the name in memory and some of these defects are simply typing errors.

We expect a large effect associated with the length of a name in characters- the longer the name, the more opportunity for both misremembering and for mistyping.

We also expect that dictionary words are more likely to be coded and typed accurately than are non-dictionary words, although of course there are some dictionary words that are often misspelled and mistyped.

We expect that when numerals are used in a name (e.g., 1source.net vs. onesource.net) that the user is more likely to miscode the name, recalling it in one form when the actual domain takes the other.

Similarly, we expect that when domain names include hyphens (e.g., 1-source.net or one-source.net) typing defects are likely, in that users may fail to accurately recall whether hyphens are used or not.

We expect that most users type *.com* as the default top-level domain when they type in a domain name, and so when a domain name uses *.net* as its top-level domain, coding defects are introduced more frequently.

By similar logic, we expect that domains that use the article “the” in their name are at risk for coding defects (e.g., “thecontainerstore.com” vs. “containerstore.com”), because some users will remember the name without the article.

Some of these factors can be mitigated by registration of multiple spelling variants of a name (e.g., registering 1-source.net, 1-source.com, 1source.net, and 1source.com). However, the typical motive for introducing hyphens, articles, and the .net top-level domain, is that the more standard configuration has already been registered. Thus, we expect that not only do these naming practices introduce coding defects, but they will sometimes result in users being directed to unintended websites.

Table 1 summarizes the attributes of names that may influence awareness, preference, and navigation yield. In some cases, an attribute may relate to more than one dimension of performance. In some cases we have clear a priori expectations about the direction of a relationship; in others we do not.

4 Data

We expect the effects described above to be subtle in comparison to the overwhelming importance of the semantics of name itself (e.g., the meaning of the words in JiffyLube.com). As a result, we are likely to need a very large sample to associate these variables with performance. We use internet domain names because we have a large sample of relatively objective, uniformly measured performance data for websites and because the names must be unique and unambiguous when encoded as a domain name.

4.1 Dependent Variables

Our dependent variable, capturing website performance, is obtained from the Fall 2009 editions of the “top million websites” reports provided by the internet data companies Alexa and Quantcast. Both companies use a similar approach to estimate the top million websites based on web traffic. They recruit a panel of several million internet users and those users install a “plug in” for their browser which records what sites they visit. The companies then use this sample to rank the websites based on the number of unique visitors. A list of the top million websites is then published. While the top million are all ranked, the traffic numbers estimated are only reported for a very small subset of high volume websites. To test our effects for both high volume and low volume websites, we set rank as our dependent variable. The reported ranks are of course estimates, and they are likely to be fairly noisy estimates. However, errors in the estimates are unlikely to be correlated with the variables in our hypothesis tests.

We believe our theory applies most to commercial websites, and we thus restrict our attention to dot-com and dot-net sites in these reports. Combining the data from two sources, we have a pooled sample of 1.28 million unique rank-domain-name pairs. About 40.5% of the rank-domain-name pairs are obtained from the Alexa ranking and the rest from Quantcast. 137,901 of the commercial domain names appear in the top 1 million lists in both datasets, albeit often with slightly different ranks.

4.2 Control Variables

We expect the visibility of a website to grow with time. Thus, older websites are likely to have a higher traffic performance and rank. To control for this effect of age, we obtained the date of domain registration through an automated query (i.e., “scraping”) of the VeriSign who-is registry from the “creation date” field, and define age as December 2009 less the creation date.

Website traffic also varies widely by the type of activity (e.g., news, sports, etc.) of the website. To control for the effect of the website category, we use data from a commercial, managed URL blacklist service (urlblacklist.com) that indexes domains by their type of activity. The database is updated daily with manual user submissions and automated inputs from a collection of internet directory websites. Many changes and new submissions are human verified. The same database is used by commercial and free internet filtering software. While the list is extensive, it does not categorize all domains in our dataset. Domains may belong to more than one category. A full list of categories is provided in Table 9 in the Appendix.

4.3 Computed Variables

We used simple string processing to compute these explanatory variables: length of the domain, .com vs. .net, use of at least one hyphen, use of “the”, and the use of numerals.

These explanatory variables are based on pronunciation: the number of syllables, the degree of alliteration, the extent of assonance, and the extent of consonance. We base pronunciation on the public domain CMU pronouncing dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict). This is a machine-readable dictionary of North American English that contains over 125,000 words and their transcription into 39 basic phonemes, along with their different variants due to lexical stress. The dictionary includes most common names, abbreviations and other proper nouns. This dictionary is extensively used in speech recognition and synthesis software. For each of our domain names, we split the domain name into the most likely combination of words contained in this dictionary and use the obtained transcription as the basis for computing our phonetic variables. The number of syllables is coded from the lexical stress transcription in the pronunciation.

For domains that contain vowel sounds, we compute assonance score as

$$\frac{Maximum\ Number\ of\ Times\ a\ Single\ Vowel\ Sound\ Appears}{Total\ Number\ of\ Vowel\ Sounds}$$

For domains with no vowel sounds, the assonance score is set to zero. The consonance score is computed in the same fashion but for consonant sounds. Finally, we count the instances of an alliteration pattern of sounds and define that as the extent of alliteration.

Finally, to compute the fraction of dictionary words used in the domain name, we extract commonly used English words from the domain. Our list of common English words comes from the public domain dictionary package 12dicts (http://wordlist.sourceforge.net/). The list contains 32,153 commonly used English words including abbreviations, acronyms, hyphenations, names and phrases. Each domain name is parsed into the most probable split of words from this dictionary, where the probability of any split is computed using the frequency of use of the constituent words in the English language. The fraction of dictionary words is computed as the ratio of the number of letters in the extracted dictionary words and the total number of letters in the domain name. Table 2 describes a record from our database.

Summary statistics of all variables in our sample, and in the constituent sub-samples are provided in Table 3. Table 4 provides the Pearson correlation coefficients between our computed variables.

5 Analysis and Results

Our dependent variable, traffic rank, only takes discrete values. Two basic approaches are proposed for such ordinal data (cf. (Piccolo, 2006) for recent developments and (Marden, 1995) for traditional approaches). The first approach typically is based on an extension of discrete choice models where the assigned ranks are considered as ordered set of categories (cf. ordered Logit model (Greene, 2003)). Typically, the choice of category is driven by the transformation of an unobserved variable into the observed ordinal categories. The unobserved variable is modeled to include covariates and some uncertainty. An alternate approach directly models the rank data using discrete probability distributions such as the Inverse Hypergeometric distribution ((D'Elia, 2003)), or a mixture of distributions ((D'Elia & Piccolo, 2005) uses a mixture of uniform and binomial distributions). Both these approaches take into account the discrete nature of the data and allow considering extreme values differently than moderate values. These approaches are most useful when there are multiple respondents that rank items into a limited number of ranks (typically less than 50).

Websites in our dataset are classified into 1 million different ranks (categories) and the advantages of using a model with discrete categories or modeling the rank data as a discrete probability distribution are very limited at this scale. Thus, we test our hypothesized relationships by estimating an OLS regression model for the website performance, as reflected by its rank. We set the log of the reported rank as the dependent variable, and the above constructed variables are used as explanatory variables. Using ranks as the dependent variable leads to identical inferences. The results of our estimation are reported in Table 5.

Our control variables are the age of the website (in years), the source of the ranking (Alexa or Quantcast) and the category of the website. Columns 1 and 2 of Table 5 show the estimates from a model that includes only the control variables. The category controls are included in Column 2. As expected, we find that older websites (higher age) are more popular (lower rank). A one standard deviation increase in a website's age improves its rank by 10.16%.[2] The effects of category are particularly salient. Adult, Advertising, E-commerce, Phishing, Gaming and News websites have higher ranks than Vacation, Medical, Religious, and Book websites.

Column 3 of Table 5 presents the estimates from a model that includes all our hypothesized effects except the role of disgust phonemes. As a whole, all of our models have explanatory power, i.e. the estimated coefficients are jointly different from zero. Our models explain between 5 and 6 percent of the variance in rank, with the full models explaining an additional 1 percent over the control variables alone.

We find that shorter domains (measured as number of characters or syllables) perform better. This suggests that, in contrast to the literature that found recall benefits of longer names in the offline world, with respect to domain names, the increased navigational yield effect of shorter names may exceed any benefits to recall of longer names. A one standard deviation increase in the length of a domain name measured as number of characters (number of syllables) is associated with a decline in site rank of 7.34% (2.99%). Alternately, an extra character or syllable is associated with a 1-2% lower rank.

Next, our estimates reveal that the use of a numeral in a domain name is associated with an improvement in the website rank of 8.19%. We had hypothesized that the use of numerals and other unconventional spellings increase recall, but may increase coding errors and limit the navigational yield. Our results suggest that on balance the use of numerals is associated with better rank.

We find that the fraction of dictionary words is associated with positive site performance. In particular, an increase in the fraction of dictionary words by one standard deviation is associated with an improvement in rank of about 1.5%. As an extreme, we predict that a completely synthetic name's rank would be about 5% worse than the rank of a name composed entirely of dictionary words.

[2] Percentage changes in ranks due to any explanatory variable are computed as $10^{s.d.(\beta)} - 1$, where x is 1 for dummy variables, and the standard deviation for continuous variables; β is the coefficient for the explanatory variable averaged over all models in which the variable was present.

We find no statistically significant association of rank with alliteration. However, both assonance and consonance are associated with worse rank. These effects, while statistically significant, are smaller than the other effects reported.

As hypothesized, we find that domain names with the .net top-level domain perform significantly worse than .com names—the rank is 4.42% worse on average. Similarly, the use of a hyphen in the name is also associated with a website rank about 2.9% worse. We find no association with the use of “the” in the domain name. Typically, “the” is used at the start of the domain names, and our regression model also includes dummy variables for starting phoneme classes. Not surprisingly, the dummy variable for the use of “the” doesn’t have enough variation independent of the variation in the starting phoneme dummies and consequently, our design can perhaps not pick up the effect of using “the” in a domain name.

We also tested for the impact of different starting phonemes. As hypothesized, we find that domain names that start with stop plosives are associated with higher traffic, an advantage in rank of 1.86% compared to names that start with lateral consonants. We also find that simple vowels (i.e., a, e, i, o, u), fricative consonants, and glides are associated with improvements in rank of 3.71%, 3.34% and 2.75% respectively (all relative to lateral consonants). We conjecture that this effect is driven in part by the unique use of the vowel sound “e” in the domain name context (such as ebay.com, ehow.com). Unlike simple vowel sounds, combination consonants are associated with a 2.27% worse rank.

The presence of disgust phonemes in the name is associated significantly with traffic. As we had hypothesized, the relationship varies greatly with the category of the website. We find that the presence of a disgust phoneme in the website’s name is associated with lower performance for gaming websites (42.73%), news websites (16.18%), pet websites (12.74%), ad websites (8.99%), religious websites (7.84%) and e-commerce websites (4.83%). On the other hand, traffic for adult and sporting websites are positively associated with the inclusion of disgust phonemes. An adult website’s rank is higher by 8.69% with the inclusion of a disgust phoneme, a sport website higher by 7.26%.

To test the robustness of our results, we ran a number of alternate specifications. Table 6 reports the results from some of the robustness tests we ran. One potential concern from our results could be systematic biases in the measurement of traffic. To address these concerns, we used two distinct independent sources of website traffic—Alexa and Quantcast. Alexa and Quantcast focus on slightly different populations of websites. While Alexa includes all websites, Quantcast excludes many sites with content deemed to be "inappropriate". Alexa's panel of web-users is more international than that of Quantcast. Finally, the algorithms used to identify unique visitors and to detect manipulation through automated tools are different.[3] To address any concerns around biases that arise from the idiosyncrasies of the measurement mechanisms, we estimate our model separately for the two sources of our dependent variable. Column 1 of the Table provides estimates from the subsample of ranks obtained from the Alexa traffic reports (518,258 websites). Column 2 does the same for the Quantcast traffic reports (758,902 websites). Columns 3 and 4 estimate a model with the average rank of the subset of websites that are ranked by both Quantcast and Alexa (137,901 websites).

Another concern with our reported effects is that the effect of morphological properties of a domain name may vary significantly between the top websites and less popular websites. The top websites may be associated with organizations that have invested in establishing a brand and the traffic could be driven by such marketing actions, beyond the properties of the name. To address this concern, we estimate our model excluding the top websites. Columns 6 and 7 exclude any websites with ranks better than 10,000 and 100,000 respectively. To the best of our knowledge none of the websites in these subsamples have any significant marketing expenses.

In each of the above described estimates, the direction of effects predicted remains the same irrespective of the sample used to estimate our model. Further, for a vast majority of the effects, the magnitude of the effect is also roughly the same. Our results are thus unlikely to arise from biases in measuring web traffic

[3] See http://bit.ly/d2o8sX for a full comparison. The seomoz.org blog provides an updated comparison of different web-traffic measurement tools.

by either of our two independent sources of traffic, and our conclusions apply equally to top websites and websites with no significant marketing efforts or skill. This suggests that it is thus unlikely that the effects are driven by any established or well–known marketing principles.

6 Discussion

We find strong evidence that the morphological and phonological properties of a domain name are associated with the level of traffic of the associated website. Here we discuss the magnitude of these effects, the limitations of the study, the competing hypotheses consistent with our results, and the implications for names in the off-line world.

6.1 Magnitude of Effects

(Brynjolfsson & Smith, 2003) shows that a log-log model can be used to estimate sales from sales rank. They use data on book sales and sales rank at Amazon.com. We conjecture that a similar model might be used to link traffic with traffic rank. Traffic estimates are provided by Alexa and Quantcast for some websites, although not provided for their complete “top million” data. We use this subset of domains for which we have both estimates of rank and traffic to estimate the impact of our effects on web traffic. We fit the data on traffic and traffic rank to a log- linear distribution: $\log_{10}(\text{traffic}) = a + b \cdot \log_{10}(\text{rank}) + e$, where e is orthogonal to log(rank) and is spherical, following the standard OLS assumptions. We obtain an estimate of 9.34 for a, and -1.05 for b, with an R-square of 99.98 %.[4] These estimates suggest that a percentage change in rank leads to roughly the same percentage change in traffic. Thus, a 5 percent change in rank, as is typical for most of our individual effects, would lead to a 5 percent change in traffic.

The difference between a good name and a bad name based on our model would comprise several effects (e.g., length, top-level domain, use of numerals), and so the difference in traffic associated with the

[4] Incidentally, our estimates of the traffic–rank model using web-traffic data are very close to the estimates of (Brynjolfsson & Smith, 2003) that used Amazon books sales data.

combined effects could be 10-20 percent. This magnitude seems highly significant to us relative to other kinds of marketing decisions organizations make.

Consider the illustrative examples in Table 7 for the hypothetical website *labnest.com*. Assume labnest.com has an Alexa rank of 100,000, corresponding to traffic of 150,636 unique visitors per year. Consider the implications of our model for these variants of the name: labnest.net, lab-nest.com, labfornest.com, labnet3.com, topnest.com, lubnest.com, and lebnest.com. On average, these names have predicted traffic levels that differ by 13,169 from the base case. (Note that in this illustration, we assume that the semantics of these names are equivalently compelling. That is, we do not consider the effect of the difference in meaning in English between "lab nest" and "top nest," which is likely to be significant depending on the function of the site.) Putting the different effects together, we would predict topnet3.com to have traffic of 164,338, while lub-for-nest.net would have traffic of only 84,872. In our opinion, differences of these magnitudes are highly significant in practice.

6.2 Limitations

The idiosyncrasies of our data impose some limitations on this study. Some of the more significant limitations follow.

- We are not able to distinguish between stand-alone websites and websites that are associated with either pre-existing organizations or organizations that communicate with their stakeholders via many different channels (e.g., Nike vs. Ebay).
- While we are able to control for broad categories of types of organizations, we are not able to distinguish between organizations whose website traffic is principally driven by inbound links and those whose traffic is driven principally by direct typing of domain names. We would expect names to vary in importance depending on the primary mechanisms by which traffic is generated.
- Many of our results are likely to hold primarily for English language sites. While the vast majority of our sites are targeted at English speakers, a few target other audiences.

While these limitations should be considered when interpreting our results, we would not expect these factors to be systematically correlated with the other variables in the study.

6.3 Competing Hypotheses

One can question whether the association we have identified is causal. Do properties of names cause success? The temporal sequence for domain names and website success is important and significant in addressing part of this question. With very few exceptions, domain names are established before the websites have traffic. So, it can not be that high traffic causes the selection of a particular type of name. However, it is still possible that both the properties of domain names and website traffic are caused by an omitted variable. Perhaps firms or individuals that are highly skilled and knowledgeable about business and marketing select names with particular properties and also achieve business success. It is not that the name itself has a direct effect, but rather that the omitted variable “marketing skill” explains both. Alternatively, perhaps firms or individuals that anticipate creating sites with high demand invest more heavily in creating names with particular properties. Certainly both of these explanations are likely to have some validity. However, even if these arguments were to explain all of our findings, they would still lead to a strong conclusion. Instead of concluding that properties of names *directly cause* business performance, the conclusion would be that successful managers collectively select names with the properties we have identified as associated with high business performance. Note that there is very little apparent consensus in the literature about the properties of good names, and our findings are not, as far as we know, widely taught or practiced. Thus, one would have to conclude that successful managers collectively act in the way we have described without explicit knowledge that they are acting this way. We believe that a more plausible explanation is that the association we observe is indeed at least partially causal; properties of the name itself directly influence business performance. The competing hypotheses seem more or less likely to be valid depending on the specific effect observed, and whether that effect is intuitive and/or widely accepted as a best practice. For example, the association between the length in characters of a domain name and website performance may relate in part to the omitted variable "marketing skill," as name length is certainly something that marketing managers worry about; but the association between the use of numerals in the name and traffic probably is not explained by marketing skill, as we do not believe that the numeral effect has previously been identified or described, and it is not particularly intuitive.

6.4 On-Line vs. Off-Line Context

Our results relate properties of domain names to demand for websites. We can not conclude that all of the effects we observe exist in other contexts (e.g., off-line consumer behavior). Some of the effects we observe are clearly directly related to the internet domain (e.g., "dot net" vs. "dot com"). For example, the length of a name may not be as important off line as it is on line. However, our theory would predict most other effects to hold off line as well as on line. Furthermore, there are very few businesses in which internet performance is not important, even if the bulk of the business is conducted off line.

7 Concluding Remarks

The detailed morphological and phonological properties of domain names are associated with significant differences in website traffic. We believe that these effects are at least partially causal; properties of names influence user behavior on the internet. The implication for managers is to choose names carefully. Of course, the top-level semantics of the name are likely to be most important. FreshFish.com is probably a better name for a seafood website than StaleCatch.com. However, once a set of names with the right associations has been generated, the nuances of those names should be considered in light of likely user behavior. The results of this study can be used as heuristics in both generating and selecting names. FreshFish.com may be better than Fresh-Fish.net. 99Fish.com may be better still.

References

Aaker, D. A. 1991. *Managing Brand Equity*. New York: Free Press.

Alter, A. L. & Oppenheimer, D. M. 2006. Predicting short-term stock fluctuations by using processing fluency. *Proceedings of the National Academy of Sciences*, 103(24): 9369.

Atkinson, R. D., Ezel, S. J., Andes, S. M., Castro, D. D. & Bennett, R. 2010. The Internet Economy 25 Years After .Com: Transforming Commerce & Life. *The Information Technology & Innovation Foundation*.

Brynjolfsson, E. & Smith, M. D. 2003. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, 49(11): 1580-1596.

Caplan, D., Rochon, E. & Waters, G. S. 1992. Articulatory and phonological determinants of word length effects in span tasks. *The Quarterly Journal of Experimental Psychology Section A*, 45(2): 177-192.

Chiravuri, A. & Peracchio, L. 2003. Investigating online consumer behavior using thin slices of usability of web sites. : 2105-2110.

Coltheart, V., Patterson, K. & Leahy, J. 1994. When a ROWS is a ROSE: Phonological effects in written word comprehension. *The Quarterly Journal of Experimental Psychology Section A*, 47(4): 917-955.

D'Elia, A. 2003. A mixture model with covariates for ranks data: some inferential developments. *Quaderni di Statistica*, 5: 1-25.

D'Elia, A. & Piccolo, D. 2005. A mixture model for preferences data analysis. *Computational Statistics \& Data Analysis*, 49(3): 917-934.

Edelman, B. 2008. Unintended Adventures in Browsing. , Fall.

Frankel, A. 2005. *Wordcraft: The Art of Turning Little Words Into Big Business*: Three Rivers Pr.

Greene, W. H. 2003. *Econometric analysis*. Upper Saddle River, NJ: Prentice Hall.

Haried, P. & Zahedi, F. 2006. Thin Slice Judgment in Predicting Web Visitors' Behavior. *AMCIS 2006 Proceedings*: 39.

Hennessey, J. E., Bell, T. S. & Kwortnik, R. J. 2005. Lexical interference in semantic processing of simple words: Implications for brand names. *Psychology and Marketing*, 22(1): 51-69.

Jakobson, R. & Waugh, L. R. 2002. *The sound shape of language*: Walter De Gruyter Inc.

Jespersen, O. 1922. Language, its nature, origin and development. *New York: Holt*.

Keller, K. L. 2003. Brand synthesis: The multidimensionality of brand knowledge. *Journal of Consumer Research*, 29(4): 595-600.

Keller, K. L. & Lehmann, D. R. 2006. Brands and branding: Research findings and future priorities. *Marketing Science*, 25(6): 740.

Klink, R. R. 2003. Creating meaningful brands: The relationship between brand name and brand mark. *Marketing Letters*, 14(3): 143-157.

Klink, R. R. 2000. Creating brand names with meaning: The use of sound symbolism. *Marketing Letters*, 11(1): 5-20.

Laxon, V., Masterson, J. & Coltheart, V. 1991. Some bodies are easier to read: The effect of consistency and regularity on children's reading. *The Quarterly Journal of Experimental Psychology Section A*, 43(4): 793-824.

Lowrey, T. M. 2007. Sounds Convey Meaning: The Implications of Phonetic Symbolism for Brand Name Construction. In T. M. Lowrey (Ed.), *Psycholinguistic phenomena in marketing communications*. Mahwah, NJ: Lawrence Erlbaum.

Lowrey, T. M., Shrum, L. & Dubitsky, T. M. 2003. The relation between brand-name linguistic characteristics and brand-name memory. *Journal of Advertising*, 32(3): 7-17.

Macnamara, J. 1984. Names for things: A study in human learning.

Marden, J. I. 1995. *Analyzing and modeling rank data*: Chapman \& Hall/CRC.

Margolis, E. & Laurence, S. 1999. *Concepts: core readings*: The MIT Press.

Minton, A. 1958. Some Popular Components of Trade Names. *American Speech*, 33(2): 17-28.

Morris, E. 2004. *From Altoids to Zima: the surprising stories behind 125 brand names*: Fireside.

Peracchio, L. A. & Luna, D. 2006. The role of thin-slice judgments in consumer psychology. *Journal of Consumer Psychology*, 16(1): 25-32.

Piccolo, D. 2006. Observed information matrix for MUB models. *Quaderni di Statistica*, 8: 33-78.

Rivkin, S. & Sutherland, F. 2004. *The making of a name: the inside story of the brands we buy*: Oxford University Press, USA.

Sebba, M. 1986. The -Ex Ending in Product Names. *American Speech*, 61(4): 318-326.

Smith, G. W. 1998. The political impact of name sounds. *Communication Monographs*, 65(2): 154-172.

Song, H. & Schwarz, N. 2009. If It's Difficult to Pronounce, It Must Be Risky. *Psychological Science*, 20(2): 135.

Stewart, K. J. 2006. How hypertext links influence consumer perceptions to build and degrade trust online. *Journal of Management Information Systems*, 23(1): 183-210.

Tehan, G. & Tolan, G. A. 2007. Word length effects in long-term memory. *Journal of Memory and Language*, 56(1): 35-48.

Vanden Bergh, B. 1990. Play and problem solving: An elusive connection. *Merrill-Palmer Quarterly*, 36(2): 261-272.

Vanden Bergh, B., Collins, J., Schultz, M. & Adler, K. 1984. Sound Advice on Brand Names. , 61: 835-840.

Yorkston, E., Menon, G. & Mick, D. G. 2004. A sound idea: Phonetic effects of brand names on consumer judgments. *Journal of Consumer Research*, 31(1): 43-51.

Tables and Figures

Figure 1: Model of role of names in website performance

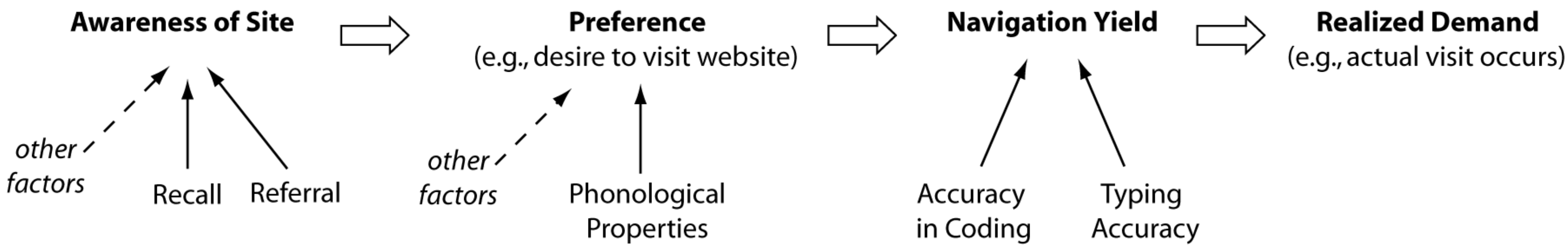

Table 1: Attributes of domain names and their hypothesized relationship to awareness, preference, navigation yield, and net effect on traffic rank.

	Awareness of Site (Recall)	Awareness of Site (Search and Referrals)	Preference for Visiting Site	Yield on Attempt to Visit Site	Net Effect on Traffic Rank
Length – characters	+			-	?
Length – syllables	-				-
Numerals in name	+			-	?
Starts with Stop Plosives	+				+
Dictionary word fraction	+	+		+	+
Alliteration	?		?		?
Consonance	?		?		?
Assonance	?		?		?
Starts with combination consonants	?		?		?
Starts with diphthong	?		?		?
Starts with fricative vowel	?		?		?
Starts with glide	?		?		?
Starts with lateral consonant	?		?		?
Starts with Nasal Consonant?	?		?		?
Starts with RC Vowel?	?		?		?
"Disgust" phonemes			- (for most categories)		- (for most categories)
Use of "the" in name				-	-
Hyphenation		+		-	-
.net instead of .com			-	-	-

Table 2: Example record for the domain name ebay.com

Name	ebay.com
Age	14.46 years
Category	e-commerce
Rank	25
Rank source	Alexa
Phonemes	EH1 B EY1
Length – characters	4
Length – syllables	2
Dictionary words	Bay
Fraction of dictionary words	0.75
Numerals in name	No
Starts with Stop Plosives	No
Alliteration	0
Consonance	0
Assonance	0
Starts with combination consonant	No
Starts with diphthong	No
Starts with fricative consonant	No
Starts with glides	No
Starts with lateral consonant	No
Starts with nasal consonant	No
Starts with RC vowel	No
Starts with simple vowel	Yes
“Disgust” phonemes	No
Use of “the” in name	No
Hyphenation	No
.net instead of .com	No

Table 3: Summary Statistics

	Combined Sample					*Domains in Alexa Top 1 Million*					*Domains in Quantcast 1 Million*				
Variable	***N***	***Mean***	***s. d.***	***Max***	***Min***	***N***	***Mean***	***s. d.***	***Max***	***Min***	***N***	***Mean***	***s. d.***	***Max***	***Min***
Age (in years)	1277166	5.86	3.97	24.86	0.07	518263	5.20	3.77	24.86	0.07	758903	6.32	4.04	24.75	0.07
Length - characters	1386890	12.03	4.75	63	1	580959	11.05	4.50	63	1	805931	12.74	4.80	63	1
Length - syllables	1386883	4.23	1.67	68	0	580953	4.05	1.63	68	0	805930	4.36	1.69	24	0
.net?	1386890	0.09	0.29	1	0	580959	0.11	0.32	1	0	805931	0.07	0.26	1	0
Hyphen?	1386890	0.09	0.29	1	0	580959	0.11	0.32	1	0	805931	0.08	0.27	1	0
The?	1386890	0.02	0.14	1	0	580959	0.01	0.12	1	0	805931	0.02	0.14	1	0
Numeral?	1386890	0.06	0.24	1	0	580959	0.08	0.28	1	0	805931	0.05	0.21	1	0
Fraction Dictionary	1386883	0.69	0.32	1	0	580953	0.69	0.31	1	0	805930	0.70	0.33	1	0
Alliteration	1386883	0.11	0.35	14	0	580953	0.12	0.36	14	0	805930	0.11	0.34	7	0
Consonance Score	1386883	0.19	0.17	1	0	580953	0.19	0.18	1	0	805930	0.20	0.16	1	0
Assonance Score	1386883	0.23	0.26	1	0	580953	0.23	0.27	1	0	805930	0.24	0.25	1	0
Disgusting Sound?	1386890	0.20	0.40	1	0	580959	0.21	0.41	1	0	805931	0.19	0.39	1	0
Combination Consonant?	1386883	0.04	0.19	1	0	580953	0.03	0.18	1	0	805930	0.04	0.19	1	0
Diphthong?	1386883	0.03	0.18	1	0	580953	0.04	0.19	1	0	805930	0.03	0.17	1	0
Fricative Consonant?	1386883	0.24	0.43	1	0	580953	0.24	0.43	1	0	805930	0.24	0.43	1	0
Glides?	1386883	0.09	0.28	1	0	580953	0.09	0.28	1	0	805930	0.09	0.28	1	0
Lateral Consonant?	1386883	0.03	0.18	1	0	580953	0.03	0.17	1	0	805930	0.03	0.18	1	0
Nasal Consonant?	1386883	0.09	0.29	1	0	580953	0.09	0.29	1	0	805930	0.10	0.29	1	0
RC Vowel?	1386883	0.00	0.06	1	0	580953	0.00	0.06	1	0	805930	0.00	0.06	1	0
Stop Plosive?	1386883	0.32	0.47	1	0	580953	0.31	0.46	1	0	805930	0.32	0.47	1	0
Simple Vowel?	1386883	0.16	0.36	1	0	580953	0.17	0.37	1	0	805930	0.15	0.36	1	0
Adult?	1386890	0.042	0.200	1	0	580959	0.074	0.262	1	0	805931	0.018	0.134	1	0
Porn?	1386890	0.015	0.120	1	0	580959	0.026	0.159	1	0	805931	0.006	0.078	1	0
Vacation?	1386890	0.011	0.106	1	0	580959	0.009	0.094	1	0	805931	0.013	0.114	1	0
E-commerce?	1386890	0.027	0.161	1	0	580959	0.018	0.133	1	0	805931	0.033	0.178	1	0
Phishing?	1386890	0.001	0.034	1	0	580959	0.002	0.041	1	0	805931	0.001	0.028	1	0
Medical?	1386890	0.009	0.096	1	0	580959	0.006	0.079	1	0	805931	0.012	0.107	1	0
Religion?	1386890	0.005	0.068	1	0	580959	0.003	0.056	1	0	805931	0.006	0.076	1	0
Sports?	1386890	0.007	0.084	1	0	580959	0.005	0.073	1	0	805931	0.008	0.091	1	0
Proxy?	1386890	0.001	0.032	1	0	580959	0.002	0.042	1	0	805931	0.000	0.022	1	0
Ads?	1386890	0.003	0.050	1	0	580959	0.003	0.054	1	0	805931	0.002	0.048	1	0
Pets?	1386890	0.002	0.047	1	0	580959	0.001	0.035	1	0	805931	0.003	0.054	1	0
Alcohol?	1386890	0.001	0.032	1	0	580959	0.001	0.027	1	0	805931	0.001	0.036	1	0
Books?	1386890	0.001	0.038	1	0	580959	0.001	0.034	1	0	805931	0.002	0.041	1	0
Magazines?	1386890	0.002	0.048	1	0	580959	0.003	0.051	1	0	805931	0.002	0.047	1	0
News?	1386890	0.003	0.054	1	0	580959	0.003	0.052	1	0	805931	0.003	0.055	1	0
Games?	1386890	0.003	0.052	1	0	580959	0.004	0.063	1	0	805931	0.002	0.042	1	0

Table 4: Pearson Correlation Coefficient (Coefficients smaller than 0.1 are not shown, All shown coefficients are statistically different from zero)

		0	1	2	3	4	5	6	7	8	9	10
Age	0											
Length- Characters	1											
Length – Syllables	2		0.721									
.net?	3											
Hyphen?	4		0.137									
the?	5											
Numeral?	6			0.121								
Fraction Dictionary	7											
Alliteration	8			0.153				0.137				
Consonance Score	9		0.176	0.211						0.363		
Assonance Score	10		0.181	0.293							0.114	

Table 5: Regression Results

Dependent Variable: Log_{10}(Rank)	Controls	Category Controls	Main Effects	Disgust Effects	%age Change in Rank^
Intercept	5.667***	5.669***	5.533***	5.534***	
Age (in years)	-0.015***	-0.013***	-0.011***	-0.011***	-10.16%
Alexa?	-0.04***	-0.028***	-0.012***	-0.012***	-3.87%
Category Controls		*Included*	*Included*	*Included*	
Morphological Attributes					
Length – Characters			0.006***	0.006***	7.34%
Length – Syllables			0.008***	0.008***	2.99%
Numeral?			-0.037***	-0.036***	-8.09%
Fraction Dictionary			-0.021***	-0.021***	-1.56%
Alliteration			0.000	0.000	
Consonance Score			0.013***	0.012***	0.67%
Assonance Score			0.01***	0.009***	0.50%
.net?			0.019***	0.019***	4.42%
Hyphen?			0.013***	0.013***	2.94%
The?			0.001	0.001	
Starting Phonemes†					
Combination Consonant?			0.01***	0.01***	2.27%
Diphthong?			-0.006*	-0.006*	-1.37%
Fricative Consonant?			-0.015***	-0.015***	-3.34%
Glides?			-0.012***	-0.012***	-2.75%
Nasal Consonant?			-0.014***	-0.014***	-3.12%
RC Vowel?			0.007	0.007	
Stop Plosive?			-0.008***	-0.008***	-1.86%
Simple Vowel			-0.016***	-0.016***	-3.71%
Disgust Phonemes					
Disgusting Phoneme?				-0.006***	-1.45%
Disgust x Adult				-0.033***	-7.24%
Disgust x Porn				-0.001	
Disgust x Vacation				-0.004	
Disgust x E-commerce				0.026***	6.28%
Disgust x Phishing				0.019	
Disgust x Medical				0.009	
Disgust x Religion				0.039**	9.29%
Disgust x Sports				-0.026*	-5.81%
Disgust x Proxy				-0.025	
Disgust x Ads				0.043*	10.44%
Disgust x Pets				0.058**	14.24%
Disgust x Alcohol				-0.001	
Disgust x Books				0.023	
Disgust x Magazines				0.005	
Disgust x News				0.071***	17.63%
Disgust x Games				0.159***	44.18%
N	1277166	1277166	1277160	1277160	
Adjusted R^2	0.0191	0.0425	0.0527	0.0529	
F Value	12424.7	3148.97	1976.36	1347.22	

†- Lateral Consonants are set as the base level for all starting phoneme dummies. ***-significant at < 0.1% level, ** -significant at 1% level, * significant at the 5% level. ^- Percentage changes in ranks due to any explanatory variable are computed as $10^{s.d.(\beta)} - 1$, where x is 1 for dummy variables, and the standard deviation for continuous variables; β is the coefficient for the explanatory variable averaged over all models in which the variable was present.

Table 6: Robustness Tests

Dependent Variable: Log$_{10}$(Rank)	Alexa	Quantcast	Average Rank	Average Rank^	Ranks >10K	Ranks >100K
Intercept	5.452***	5.6***	5.258***	4.786***	5.574***	5.693***
Age (in years)	-0.007***	-0.014***	-0.009***	-0.007***	-0.009***	-0.006***
Alexa?					-0.014***	0.003***
Category Controls	***Included***	***Included***	Included	Included	Included	Included
Morphological Attributes						
Length – Characters	0.01***	0.003***	0.009***	0.006***	0.005***	-0.011***
Length – Syllables	0.003***	0.012***	0.013***	0.012***	0.006***	0.002***
Numeral?	-0.016***	-0.05***	-0.01	-0.009	-0.036***	0
Fraction Dictionary	-0.022***	-0.03***	-0.025***	-0.017***	-0.021***	-0.02***
Alliteration	0.003	-0.002	-0.002	-0.003	0.001	-0.014***
Consonance Score	0.003	0.023***	0.014	0.013*	0.006**	0.002**
Assonance Score	0.009***	0.007***	0.004	0.004	0.007***	0
.net?	0.018***	0.023***	0.05***	0.043***	0.011***	0.002***
Hyphen?	0.038***	-0.01***	0.041***	0.033***	0.006***	0.001
The?	0.003	-0.002	0.004	0.003	-0.001	0
Starting Phonemes†						
Combination Consonant?	0.004***	0.013***	0.014	0.008	0.008**	0.005**
Diphthong?	0.004***	-0.012***	-0.015	-0.015*	-0.004	-0.001
Fricative Consonant?	-0.009	-0.016***	-0.044***	-0.037***	-0.011***	-0.005***
Glides?	-0.005	-0.015***	-0.092***	-0.078***	-0.005*	0.001
Nasal Consonant?	-0.012*	-0.013***	-0.024**	-0.021***	-0.011***	-0.006***
RC Vowel?	0.018	-0.001	0.021	0.016	0.005	0.001
Stop Plosive?	-0.003*	-0.01***	-0.031***	-0.026***	-0.005**	-0.001
Simple Vowel	-0.008**	-0.023***	-0.023**	-0.019***	-0.013***	-0.006***
Disgust Phonemes						
Disgusting Phoneme?	0.000	-0.011***	0.000	0.000	-0.005***	-0.002**
Disgust x Adult	-0.042***	-0.027*	-0.047*	-0.026	-0.031***	-0.013***
Disgust x Porn	0.01	-0.019	-0.047	-0.04	0.011	0.006
Disgust x Vacation	0.013	-0.016	0.013	0.013	-0.01	-0.01
Disgust x E-commerce	0.021	0.029***	0.03*	0.026*	0.021***	0.009*
Disgust x Phishing	0.008	0.038	0.04	0.028	0.015	0.02
Disgust x Medical	-0.007	0.018	-0.002	0.006	0.004	-0.002
Disgust x Religion	0.052	0.034*	0.061	0.052	0.025	0.019*
Disgust x Sports	-0.031	-0.025	-0.045	-0.034	-0.02	-0.001
Disgust x Proxy	0.013	-0.112	-0.144	-0.122	0.002	-0.009
Disgust x Ads	0.085**	0.004	0.056	0.04	0.026	0.022
Disgust x Pets	0.136**	0.051*	0.113	0.085	0.041*	-0.004
Disgust x Alcohol	0.022	0.011	0.037	0.016	0.002	-0.014
Disgust x Books	-0.034	0.061*	-0.023	-0.019	0.008	0.012
Disgust x Magazines	0.043	-0.019	-0.006	-0.009	0	-0.004
Disgust x News	0.085***	0.075***	0.07*	0.053*	0.037*	0.024*
Disgust x Games	0.129***	0.152***	0.089*	0.073*	0.087***	0.018
N	518258	758902	137901	137901	1263130	1146999
Adjusted R^2	0.046	0.0635	0.0599	0.0587	0.0411	0.0269
F Value	481.77	991.11	170.33	166.29	1022.9	600.01

†- Lateral Consonants are set as the base level for all starting phoneme dummies. ***-significant at < 0.1% level, ** -significant at 1% level, * significant at the 5% level. ^- Ranks are reassigned in sample of common websites.

Table 7: Illustrative example showing implications of model for a hypothetical gaming website labnest.com

Hypothetical Name	Notes	Estimated Rank	Estimated Traffic (unique-visitors/year)	Change in Traffic from Base Case
labnest.com		100,000 (assumed baseline)	150,636	
labnest.net	.net instead of .com	104,419	141,079	-9,557
lab-nest.com	Add hyphen	102,941	143,206	-7,430
labfornest.com	Increase length to 10.	104,585	140,844	-9,792
labnet3.com	Numeral in name	91,986	161,167	+10,530
topnest.com	Initial stop plosive "T"	98,161	150,537	-99
lubnest.com	Disgust phoneme "uh"	142,088	102,092	-48,544
lebnest.com	Fraction dictionary reduced to 4/7.	102,125	144,408	-6,228
topnet3.com	Best combination	90,294	164,338	+13,702
lub-for-nest.net	Worst combination	169,422	84,872	-65,764

This page is intentionally blank to separate the paper from its Appendix.

Appendix

1 Phoneme Categories

Table 8: Basic Phoneme Categories and the included Phonemes

Phoneme Category	Phonemes (IPA symbol, Example)	
Stop Plosive Consonants (SP)	P	***p***ay
	B	***b***uy
	T	***T***ake
	D	***d***ay
	K	***k***ey
	g	***g***o
Nasal Consonants (NC)	M	***m***an
	N	***n***o
	Ŋ	si***ng***
Fricative Consonants (FC)	F	***f***or
	V	***v***ery
	Θ	***th***anks
	Đ	***th***at
	S	***s***ay
	Z	***z***oo
	ʃ	***sh***ow
	ʒ	mea***s***ure
	H	***h***ouse
Lateral Consonants (LC)	L	***l***ate
Glides (GL)	R	***r***un
	J	***y***es
	W	***w***ay
Combination Consonants (CC)	tʃ	***ch***air
	dʒ	***g***ym (JH IH1 M)
Simple Vowel (SV)	ɔ	***o***ff
	ɑ:	b***ou***ght
	i:	b***ee***
	u:	y***ou***
	ɛ	r***e***d
	ı	b***i***g
	ʊ	sh***oul***d
	ə	b***u***t
	Æ	***a***t
Diphthongs (DV)	eı	s***ay***
	aı	m***y***
	oʊ	sh***ow***
	aʊ	h***ow***
	ɔı	B***oy***
RC Vowels (RCV)	ɝ:	H***er***

2 Categories for Classification of Domains

Table 9: Domain Categories

Adult
Porn
Vacation
E-commerce
Phishing
Medical
Religion
Sports
Proxy
Ads
Pets
Alcohol
Books
Magazines
News
Games

Europe Campus
Boulevard de Constance
77305 Fontainebleau Cedex, France
Tel: +33 (0)1 60 72 40 00
Fax: +33 (0)1 60 74 55 00/01

Asia Campus
1 Ayer Rajah Avenue, Singapore 138676
Tel: +65 67 99 53 88
Fax: +65 67 99 53 99

Abu Dhabi Campus
Muroor Road - Street No 4
P.O. Box 48049
Abu Dhabi, United Arab Emirates
Tel: +971 2 651 5200
Fax: +971 2 443 9461

www.insead.edu

INSEAD

The Business School
for the World®

Printed by INSEAD