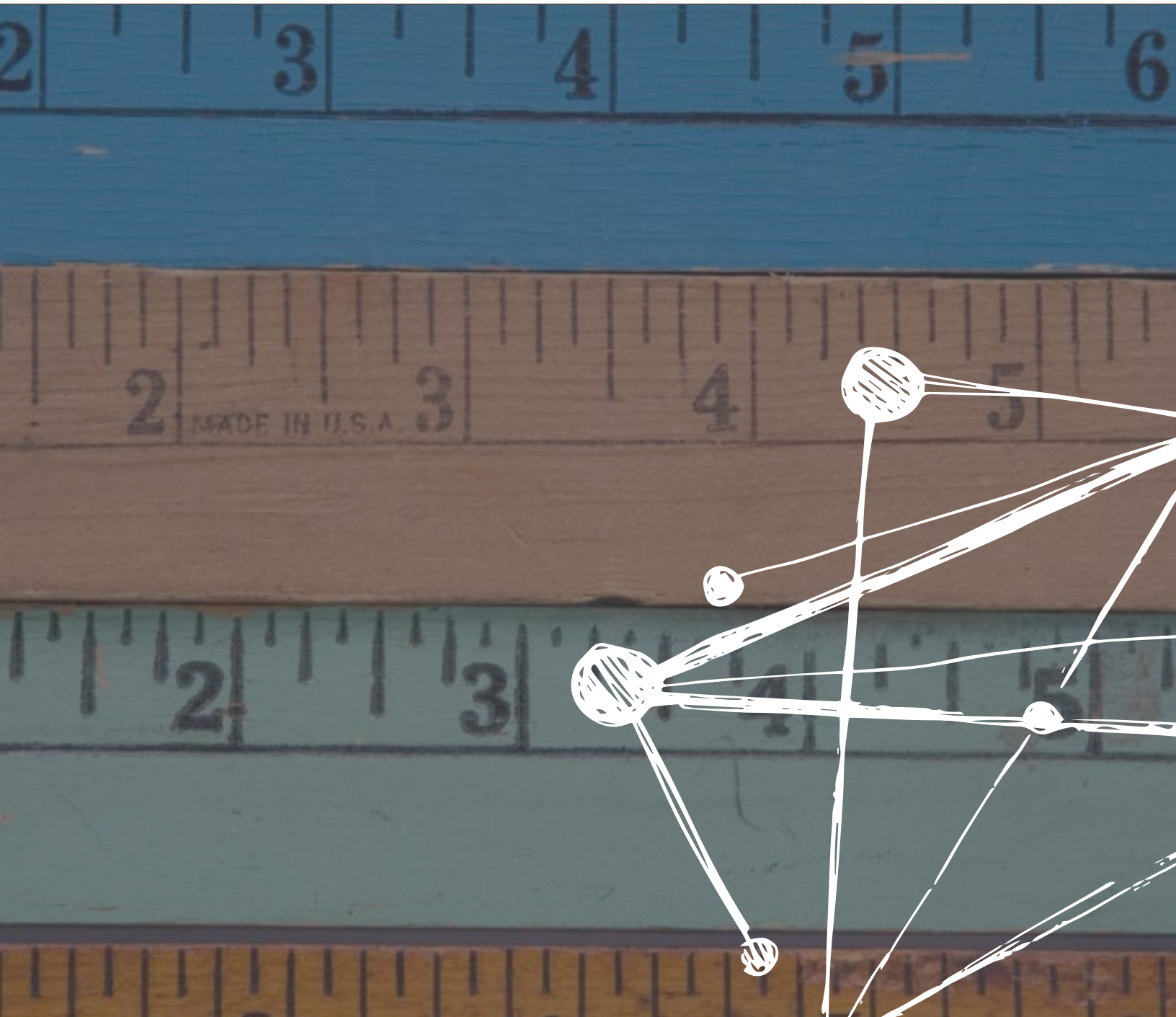
seer

The Effect of Domain Extensions on Successful Recall

Written by: Matt McAnear, Seer Alum
February 4, 2015



Abstract

Recently, new top-level domain extensions have been created that allow web-masters more customizability in their URL structures, with such endings as ".io," ".rs," and ".xyz," to name a few. Although these top-level domains allow the creation of shorter domains that were previously taken by .coms, there is a potential tradeoff in memory recall. Users may remember domains ending in .com better despite a longer domain name simply because .com is the *de facto* default top-level domain.

We conducted an experiment to test this hypothesis and analyzed the results using both a traditional logistic regression and a mixed methods logistic regression framework. We found that the total length of the domain, not a particular extension, was the most significant predictor of whether or not a participant correctly recalled a particular domain. We have identified areas for future improvement in the study.

Methodology

To begin, we created a list of potential domains from short custom functions. These functions were built using the default random number generator in R, the open-source software package and programming language used for the entirety of this paper (R Development Core Team, 2011). There were two domain creation functions, one for one-word domains ("blue.io") and multi-word domains ("big-computer-technology.tech"). The words for the beginning of the domain (everything before the ".") came from a Scrabble dictionary (Barbone, 2013).

The example output of those functions is below. Though this does create random domains for testing, we found that several were nonsensical or humorous and thus, we used the output of these functions to generate lists from which to cull our domains. This is a somewhat manual process that inevitably introduces bias into our study; a new method is recommended for any future iterations.

```
get.domain()
## [1] "answers.works"
get.long_domain("-")
## [1] "freedom-luck-background.bar"
```

After running each of these functions repeatedly (using the replicate function), we settled on eight domains. This seemed to be a optimal tradeoff among respondents during internal testing between the difficulty of the overall task and recall accuracy—attempts with more words proved to be overwhelming for participants. Here is the list of domains that were ultimately chosen.

```
## [1] "maple.com"                  "bulletin-us.xyz"
## [3] "deliveredcareful.com"       "discs.com"
## [5] "heavy.ink"                  "make.rs"
## [7] "realestatepostings.biz"     "whose-documents.com"
```

This list contains four .coms, four exotic extensions, and domain-level differences split equally between .coms and exotic domains, as follows:

- Two short and two long domains with exotic extensions
- Two short and two long .com domains
- One hyphenated long domain with an exotic extension
- One hyphenated long .com domain

From this list of eight words, we generated four different lists, the maximum number of variations in our testing tool, to control for variations in word order. Our testing solution was the 5-Second Memory Test at www.verifyapp.com. During the test, users are directed at random to one of the four lists and given five seconds to study the list. After five seconds are up, users are prompted to enter as many words as they are able into five text boxes.

Noticeably, there are fewer text boxes than words available for testing. However, on average, users correctly recalled 1.23 domains. Six users recalled five correctly, and so in six cases, this limitation may have impacted the results. However, we had 200 test subjects, and so these six users were a relatively small component of the test. Our users were acquired through verifyapp.com, a user-testing site.

Participants' responses were collected over a period of three days. Once collected, data were converted to a binary response format—each line consisted of a domain, whether or not that guess was successful, along with information both about the user and the domain. Here are ten observations from our data (not every variable shown).

```
##          domain                      correct      respondent     variation
## 1219     deliveredcareful.com        0            6121           A
## 867      deliveredcareful.com        0            5325           Control
## 1053     discs.com                   0            6892           A
## 1554     whose-documents.com         0            8087           A
## 507      deliveredcareful.com        0            4892           C
```

Analysis

Several dummy variables were added to the data to indicate whether a domain was short, had an exotic extension, and if it was the first domain of a particular variation. In our experiment, a domain was considered "short" if the word before the extension had fewer than five characters. The remaining domains were considered long, and each had at least eleven characters (including the hyphen). Initially, we also included a term that stated the exact position of a word in the list. However, only the domain in the top position was significantly more likely to be remembered, and so we included an indicator variable only for whether or not a word was at the top of the list to control for list order.

From there, we created several logistic regression models to test the significance of these factors. The first model tests the demographic factors for users, the second is domain-specific factors, and the third is the full model, which includes both demographic and domain-specific factors.

Compared to the full model, the domain specific factors resulted in the largest gains in predictive power. The demographic factors model had significant results, but did not explain as much of the variation in participants' responses as the domain-specific model. To improve the models further, a random effect was added for each respondent to estimate the intercept, using the lme4 package (Bates, Maechler, Bolker, and Walker, 2014). The code for each logistic model is below.

```
require(lme4)
logistic_1 <- glm(correct ~ age + gender + variation,
           data=df, family=binomial(link="logit"))
logistic_2 <- glm(correct ~ short*exotic + top,
           data=df, family=binomial(link="logit"))
logistic_3 <- glm(correct ~ age + gender + variation +
           short*exotic + top,
           data=df, family=binomial(link="logit"))
```

PROPORTION OF RESPONDENTS WHO CORRECTLY RECALLED A DOMAIN

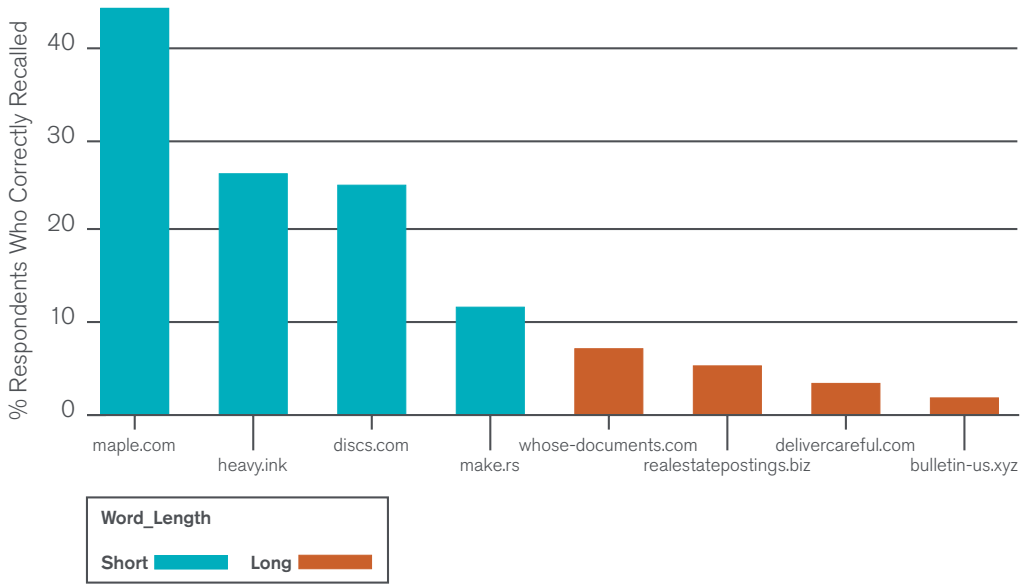


Figure 1: Distribution of correct answers by domain.

```
# Random Effects Model
logistic_mixed <- glmer(
      formula = correct ~ top + short + exotic +
            short*exotic + (1|respondent),
      family = binomial,
      data = df)
```

The results of the domain-specific model, the full model, and the mixed effects model are in the appendix for comparison.

Though age was a significant predictor of whether or not someone correctly recalled a word in previous logistic models, we decided not to include it in the mixed effects model. Once age was included in the mixed-effects model, the model no longer converged during the maximum likelihood estimation routine. Therefore, we excluded it in an attempt to get more parsimonious estimates of our main research factors. The receiver operating curve in Figure 2 is for comparison of the models, and the code is borrowed liberally from Haine (2013).

The predictive power of the mixed effects model is increased strongly by the use of the random effect as well. Adding the mixed-effects terms resulted in a model that is significantly better at predicting successes and more resilient against false positives. Because this model is significantly better than the previous logistic models, it is the one from which we will report our point estimates and confidence intervals.

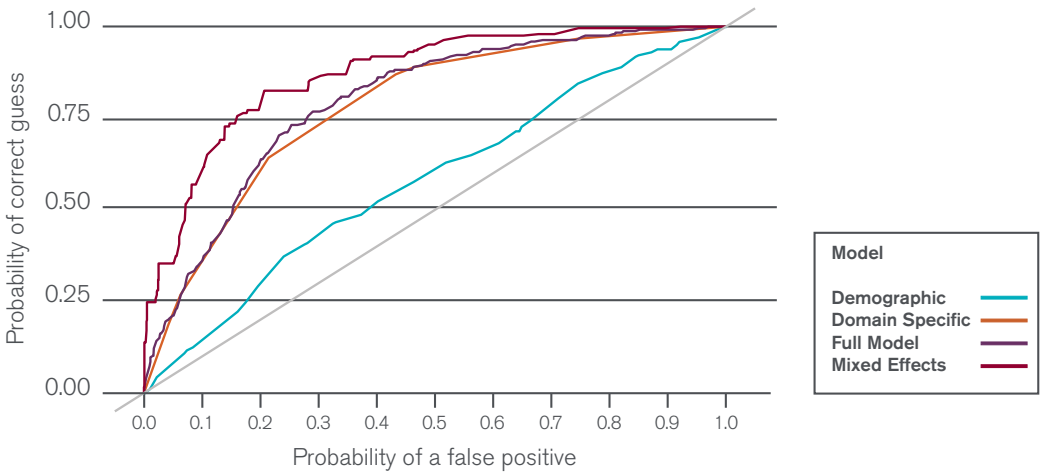


Figure 2: ROC Curves, including Mixed Effects Logistic Model.

Results

Users were overwhelmingly more likely to correctly recall a domain if it was short. Short domains were, on average, 9.88 times more likely to be recalled correctly than long domains, controlling for individual respondent and whether or not the domain was at the top of the list.

Whether or not a domain had an exotic extension was not a significant predictor of how well a participant remembered a domain. However, the p-value for this was 0.11, and thus approaches significance at the 10% level. This indicates that memory recall on domains with exotic extensions, while inconclusive in this experiment, may warrant further study.

Interestingly, the interaction term between exotic extensions and shortness was highly insignificant. While we cannot say with certainty that .coms and exotic extensions receive the same benefit from being short, the evidence against this position is very weak.

Limitations

Several experimental limitations should be addressed for future iterations of the present study. The five-second memory test employed by the project is but an approximation of memory recall ability. Further, there is the very real possibility that the domains chosen may have biased our subjects in some way. Because a maximum of four variations were allowed by the testing tool, we chose to use fewer words and have more variation in the order of the words in the lists, rather than increase the number of words but have less variation in the order. A potential improvement to the design would be to create thirty-two new domains and randomly order eight of them on a list, making sure that each had a representative URL syntax similar to the list we used.

CONFIDENCE INTERVALS: MIXED EFFECTS

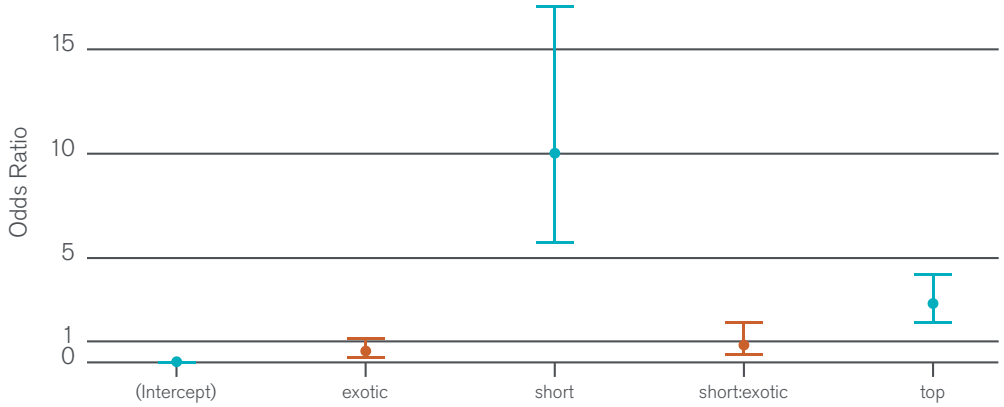


Figure 3: CI Plot: Significant Effects in Blue.

In addition, a larger sample size would aid in estimating relatively rare events, such as correctly guessing the long domain 'realestatepostings.biz.' Our estimated probabilities of correctly recalling these domains would be more accurate with a larger sample size, and this would only improve our statistical models.

Conclusion

Despite the limitations listed above, we believe that this study provides ample evidence that the length of a URL is one of the most important factors in whether or not a person correctly recalls it.

For those buying new domains, maximizing memory recall is likely a goal of the purchase. The theory behind this action is that maximizing recall will help customers who see or hear a website indirectly better remember it in the future. Our research indicates that the surest way to do this is not necessarily to focus on the extension, but instead on the length. Given two domains, a .com and one with an exotic extension, it is unclear which one a potential buyer should choose. But given the choice between a short domain and a long domain, one should choose the short domain, regardless of its extension.

Appendix and Acknowledgments

I would like to thank Seer Interactive for its generous support and financing of this study, as well as Seer Interactive employees for their participation in the initial prototype experiment.

This document was created using knitr (Xie, 2013). All figures were made using the ggplot2 package (Wickham, 2009).

Table 1: Logistic Models. Reported Coefficients are odds ratios. Confidence intervals in Mixed Effects models are bootstrapped. For the mixed effects model, stars indicate that 1 lies outside the confidence interval, not the level of significance. Table created using 'texreg' package (Leifeld, 2013).

	Domain-Specific	Full Model	Mixed Effects
(Intercept)	0:05***	0:04***	0:04*
	(0:24)	(0:49)	[0:02; 0:06]
short	8:22***	8:38***	9:88*
	(0:26)	(0:27)	[5:87; 17:48]
Exotic	0:57	0:56	0:55
	(0:37)	(0:37)	[0:25; 1:14]
top	2:46***	2:57***	2:82*
	(0:19)	(0:19)	[1:90; 4:20]
short:exotic	0:87	0:89	0:83
	(0:41)	(0:41)	[0:37; 1:94]
age21-25		2:47*	
		(0:41)	
age25-30		2:20	
		(0:41)	
age30-40		1:60	
		(0:41)	
age40-55		1:03	
		(0:45)	
age55-65		1:84	
		(0:65)	
genderMale		0:78	
		(0:17)	
variationA		0:81	
		(0:22)	
variationB		0:79	
		(0:23)	
variationC		1:06	
		(0:21)	
AIC	1159.13	1157.23	1138.40
BIC	1186.02	1232.52	1170.66
Log Likelihood	-574.56	-564.61	-563.20
Deviance	1149.13	1129.23	1126.40
Num. obs.	1600	1600	1600
Num. groups: respondent			200
Variance: respondent.(Intercept)			0.73
Variance: Residual			1.00

***$p < 0:001$, **$p < 0:01$, *$p < 0:05$ (or 1 outside the confidence interval).

References

1. Barbone, V. A. G. (2013). E-Assessment of Free-Text Answers Based on Domain Specific Sublanguages and Knowledge Representation. Universidad de la Republica, Uruguay. Retrieved from http://iie.fing.edu.uy/ vagonbar/tesis/lexicon/sources/ kevinat/english-words.10
2. Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. Journal of Statistical Software. Retrieved from http:// cran.r-project.org/web/packages/lme4/lme4.pdf.
3. Boettiger, C. (2014). knitcitations: Citations for knitr markdown les. Retrieved September 10, 2014, from https://github.com/cboettig/knitcitations.
4. Haine, D. (2013). Veterinary Epidemiologic Research: GLM Evaluating Logistic Regression Models (part 3) | (R news & tutorials). R-Bloggers. Retrieved August 09, 2014, from http://www.r-bloggers.com/veterinary-epidemiologic-research-glm-evaluating-logistic-regression-models-part-3/.
5. Leifeld, P. (2013). texreg: Conversion of Statistical Model Output in R to LATEX and HTML Tables. Journal of Statistical Software. Forthcoming, 55(8), 124. Retrieved from http://www.jstatsoft.org/v55/i08/.
6. R Development Core Team, R. (2011). R: A Language and Environment for Statistical Computing. (R. D. C. Team, Ed.)R Foundation for Statistical Computing. R Foundation for Statistical Computing. doi:10.1007/978-3-540-74686-7.
7. Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: visualizing classier performance in R. Bioinformatics, 21(20), 7881. Retrieved from http://rocr.bioinf.mpi-sb.mpg.de
8. Wickham, H. (2009). ggplot2: elegant graphics for data analysis. Springer New York. Retrieved from http://had.co.nz/ggplot2/book.
9. Xie, Y. (2013). Dynamic Documents with R and knitr. Boca Raton: CRC Press, Taylor & Francis.

PHILADELPHIA (HQ) • SAN DIEGO
p: (215) 967-4461• f: (215) 873-0744