

Tagging with del.icio.us:

Emilee Rader & Rick Wash
(ejrader,rwash@umich.edu)

University of Michigan
School of Information

Social or Selfish?

(1) Study Background

Do users of the **social bookmarking** website del.icio.us tag **socially**, choosing the same tags as others have used for a webpage? Or do they tag **selfishly**, choosing words **idiosyncratically**, resulting in the vocabulary problem identified by Furnas et al. (1983)*?

If users of del.icio.us could agree on a set of tags to use to describe a webpage, this **convergence** could make **information discovery** (finding new and interesting webpages) using del.icio.us easier. Alternatively, if the same words are used inconsistently to tag many different webpages, searching using those words would return "noisy" results.

Furnas et al. found that the probability that two people will choose the same label for an object or document is **less than 20%**. We compare the statistics they reported with those we calculated for 349 webpages bookmarked on del.icio.us, collected in December 2005.

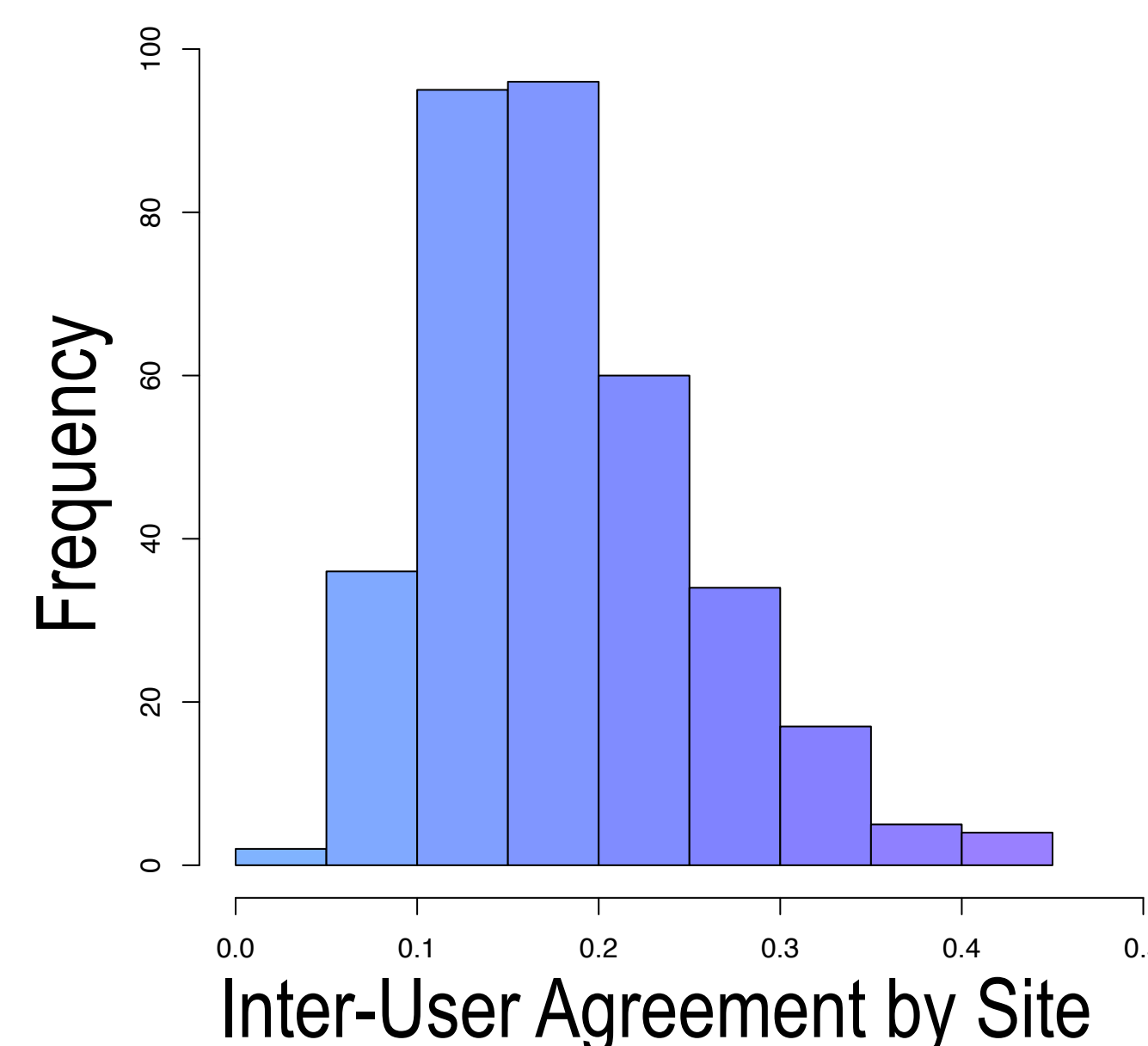
(2) Inter-User Agreement Statistics

Furnas et al.

Two random users will choose the same word less than **20%** of the time, on average.

del.icio.us

Two random users will choose the same word to tag a webpage, on average, only **17%** of the time.



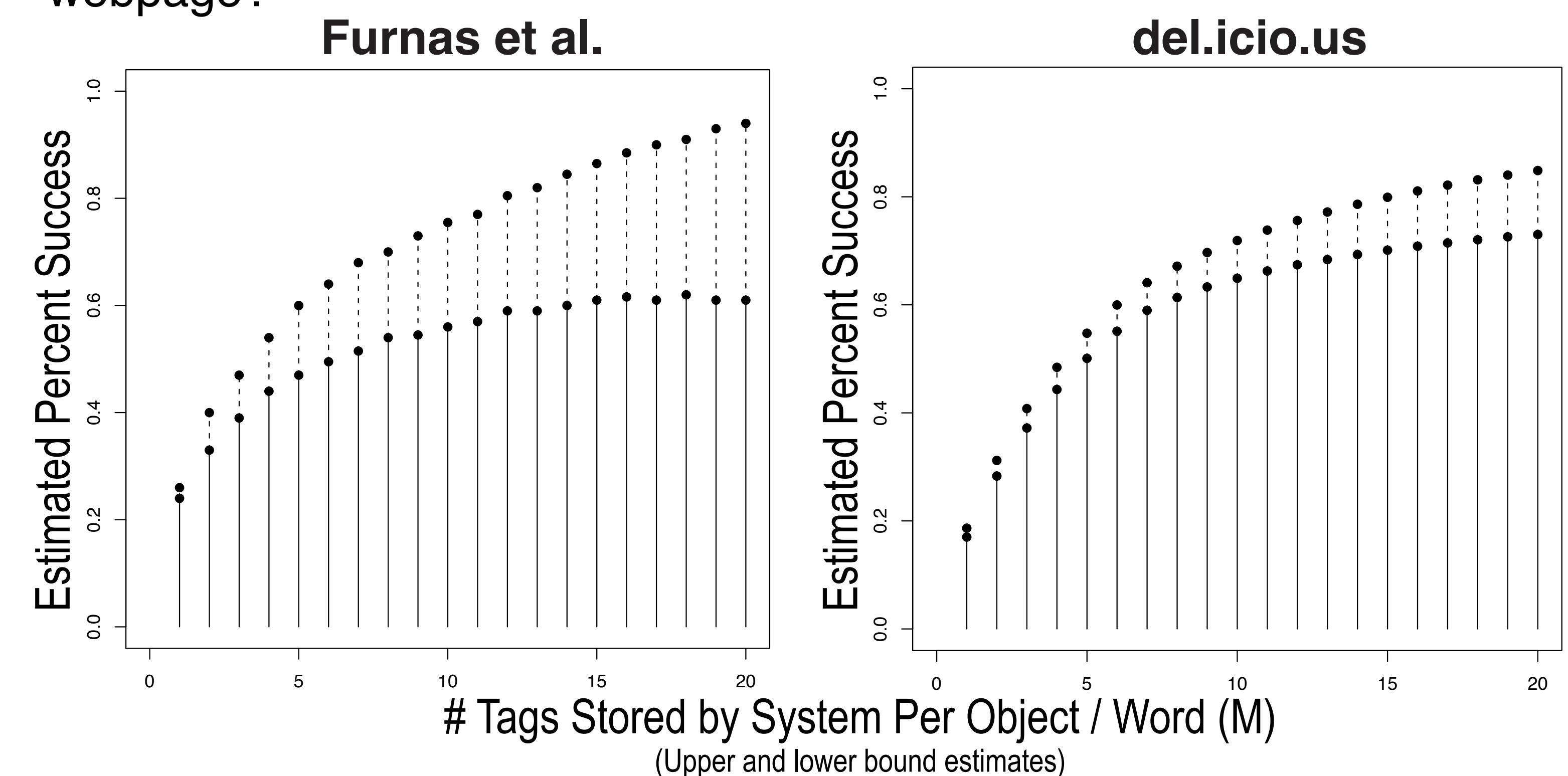
(3) Probability of Search Success

Assuming that search success depends on whether the tag by which a user chooses to search for a webpage matches one of the tags already assigned to that webpage:

1. If del.icio.us stored the **first M tags applied** to a webpage, what would be the likelihood of a search by a single tag returning that webpage?

	Furnas et al.	del.icio.us
M=1	0.12	0.08
M=2	0.21	0.15
M=3	0.28	0.22

2. If del.icio.us stored the **M tags most frequently applied** to a webpage, what would be the likelihood of a search by a single tag returning that webpage?



(4) Conclusions

Our statistics closely match those presented in Furnas et al. (1983), leading to the conclusion that the vocabulary problem exists in del.icio.us. **In general, users practice selfish tagging.**

Webpage	Saved by n users	Unique Tags	Inter-User Agreement	Top 5 Most Common Tags
Slashdot	22,193	2,598	0.14	news, technology, tech, geek, system:unfiled
del.icio.us	3,096	1,177	0.32	system:unfiled, del.icio.us, delicious, bookmarks, test
MySpace	1,482	421	0.14	system:unfiled, mspace, social, community, networking
YouTube	2,705	783	0.10	video, videos, sharing, web2.0, community
cnn.com	8,261	1,250	0.32	news, system:unfiled, cnn, media, daily
gmail	7,469	1,274	0.15	email, google, gmail, system:unfiled, mail

* Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1983). Statistical Semantics: Analysis of the Potential Performance of Key-Word Information Systems. The Bell System Technical Journal, 62(6), 1753-1806.

