

Make Data Useful

Greg Linden

amazon.com®

Findory

Introduction

- What is the goal?
- Good algorithms versus big data
- Set expectations
- Help people discover useful new stuff quickly

What is the goal?

- For what are you optimizing?

Revenue

Clicks

Visitors

Traffic

Retention

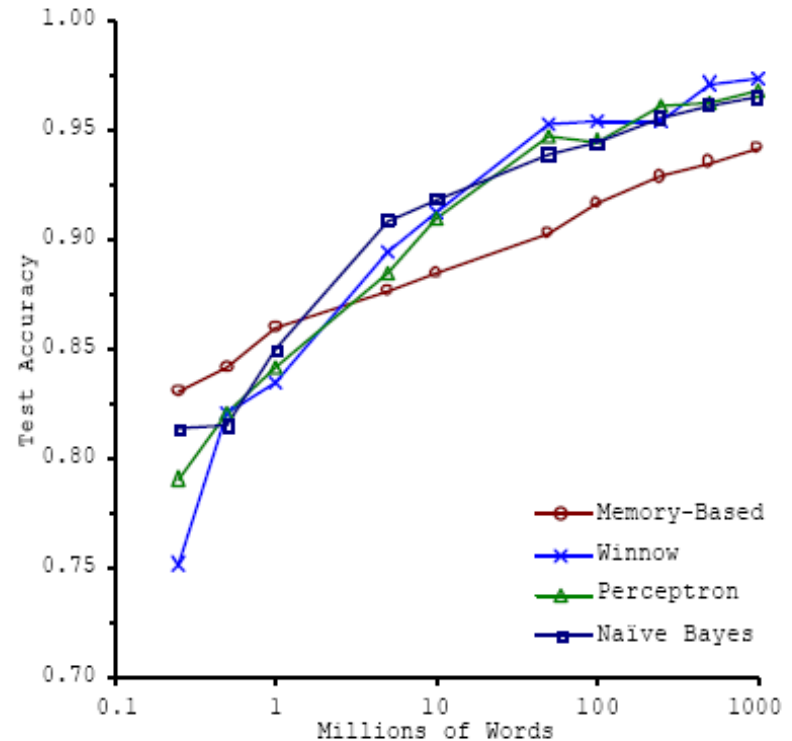
Time on site

Customer acquisition

- Measuring and optimizing
 - A/B tests
 - Measure, learn, improve, iterate
 - *“Encourage experimentation ... as much of it as possible”* - Jeff Bezos

Algorithms versus big data

“Worry about the data first before you worry about the algorithm.”
- Peter Norvig



[Banko and Brill, 2001]

Set expectations

- False positives happen
 - Design should encourage users to forgive

Customers who bought this item...



[Foundations of Statistical Natural Language Processing \(Hardcover\)](#)

by [Christopher D. Manning](#), [Hinrich Schütze](#)

Average Customer Review: ★★★★★ (10)

In Stock

This item ships for **FREE** with **Super Saver Shipping**. [Details](#)

Review

*Statistical natural-language processing is, in my estimation, one of the most fast-moving and exciting areas of computer science these days. Anyone who wants to learn this field would be well advised to get this book. For that matter, the same goes for anyone who is already in the field. I know... [Read More](#)

\$66.40

[Add to cart](#)

[Add to Wish List](#)

[40 used & new](#) from **\$58.80**

Also bought these items...

Show items from:

▶ **All Categories**

[Books](#)

Show items that:

▶ **Customers also bought**

[Customers also viewed](#)



[Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition](#) Paperback by Daniel Jurafsky

[More like this](#)



[Modern Information Retrieval](#) Paperback by Ricardo Baeza-Yates

[More like this](#)



[Mining the Web: Analysis of Hypertext and Semi Structured Data \(The Morgan Kaufmann Series in Data Management Systems\)](#) Hardcover by Soumen Chakrabarti

[More like this](#)

Be useful

- Do not get in the way
- Help when help is wanted
 - Similar books on detail pages
 - Shopping cart recommendations

Customers who bought the items in your shopping cart also bought:

- [Sid Meier's Civilization IV: Warlords Expansion Pack](#) by 2K Games
- [Microsoft Age of Empires 3](#) by Microsoft Software
- [Battlefield 2](#) by Electronic Arts
- [Quake 4](#) by Activision
- [Quake 4: Special Edition](#) by Activision

- Focus the home page content on discovery
 - A biased view of the Amazon catalog

Focus on discovery

- Recs should be non-obvious
- Recs should be hard to find yourself
 - Strike against content-based techniques?

“Discovery is when something wonderful that you didn't know existed, or didn't know how to ask for, finds you.” - Fortune Magazine

Bias toward new items

- People probably have not seen them yet
- Automated alerts, an implicit search
- People love it
 - Fresh, helpful, surprising, useful

Bias toward recent history

- Focus on the current mission
 - What people are doing now
 - What people need help with right now
- Recency matters
 - Strike against profile-based techniques?

Speed matters

- Every 100ms delay costs 1% of sales
- Performance and scalability
 - Hard with big data!
 - Needs to be part of initial architecture
 - Rules out some algorithms

Summary

- Measure, then optimize toward a goal
- Big data first, then algorithms
- Encourage users to forgive you
- Do not get in the way
- Surprise, aid discovery, focus on the new
- Speed matters