Is there a data barrier to entry?

Hal Varian
June 2015
Using robots to block Google News

We understand that news organizations publish lots of content and not all of it may be right for Google News. Google News crawls with the same robot as Google Web Search, called Googlebot.

Google Search and Google News support two different 'bots', namely Googlebot and Googlebot-News, that you can use as meta tags or in your robots entry to control where your content appears.

In other words:

If you block access to Googlebot-News, your content won't appear in Google News.
If you block access to Googlebot, your content won't appear in Google News or Web Search.
1. The concept of “data barrier to entry” dates back at least to 2007, but what does it mean?
   a. Data alone is nothing, what matters is what you do with it
2. Use and abuse of “network effects”
   a. Demand side and supply side returns to scale
   b. Demand side is not relevant for search
   c. Every successful company uses data
   d. Data is subject to diminishing returns to scale
3. Example: online search
   a. How much data is “enough”? 
   b. Building a search engine on the cheap
   c. Examples from ad targeting
4. Learning by doing and productivity growth
Economies of scale

**Demand side.** The value of adopting a service to an incremental user is larger when more users have already adopted. *Direct* and *indirect* network effects.

**Supply side.** *Scale:* The cost of producing an incremental unit is smaller at higher levels of output. *Scope:* the cost of producing an incremental unit is smaller when other related production takes place.
\[ x = \text{scale of operation} \]
\[ \text{mv}(x) = \text{value to a marginal user increases with } x \]
\[ \text{mc}(x) = \text{cost of a marginal unit produced decreases with } x \]

Consider Facebook which could conceivably have both demand-side and supply-side economies of scale

- **Demand side.** If there are more users on Facebook than MySpace, a new user would prefer to adopt Facebook.
- **Supply side.** If there are more users on Facebook than on MySpace, the average cost per user of providing the service will be lower on Facebook.
Share and scale

- **Share** is relevant for adoption decisions, **size** is relevant for cost
  - Pure network effects means bigger network is more attractive to users
  - Pure economies of scale means bigger network has lower unit cost to firm
- Don’t have to be the *most* profitable producer to survive, you just have to be profitable (i.e., cover costs)
- Upsets happen (MySpace/Facebook, Google/Yahoo/etc)
- Diseconomies of scale with respect to scale
  - Congestion
  - Competing priorities from core business needs
    - Microsoft prioritizes Windows/Office, Bing is secondary
    - Google prioritizes Search/Ads, Docs is secondary
    - A “me too” approach is futile, differentiation is key
    - Consumers benefit from competition...
Virtuous circle?

- More users
- More chances to show ads
- More search data, leading to more targeted ads
- More advertisers
- Higher ad revenue, more ads sold
- More revenue to invest in infrastructure, syndication
Economies of scale?

**Google economies of scale:**
- More users → More chances to show ads → More search data, leading to more targeted ads → More advertisers → Higher ad revenue, more ads sold → More revenue to invest in infrastructure, syndication → More users

**Magazine business:**
- More readers → More chances to show ads → More data leading to more targeted ads → More advertisers → More revenue to invest in publication → More revenue, more ad sold → More readers
Economies of scale?

**Google Confidential and Proprietary**
From virtuous circle to nutritious circle

- More users
- More revenue to invest in infrastructure, syndication
- Higher ad revenue, more ads sold
- More advertisers
- More chances to show ads
- More search data, leading to more targeted ads

Hamburger business
“The higher the number of advertisers using an online search advertising service, the higher the revenue of the general search engine platform; revenue which can be reinvested in the maintenance and improvement of the general search service so as to attract more users.”
“The higher the number of advertisers using an online search advertising service, the higher the revenue of the general search engine platform; revenue which can be reinvested in the maintenance and improvement of the general search service so as to attract more users.”
“The higher the number of advertisers using an online search advertising service, the higher the revenue of the general search engine platform; revenue which can be reinvested in the maintenance and improvement of the general search service so as to attract more users.”

“The higher the number of customers a business has, the higher the revenue of the business, revenue which can be reinvested in the maintenance and improvement of the business so as to attract more customers.”
“The higher the number of customers a business has, the higher the revenue of the business, revenue which can be reinvested in the maintenance and improvement of the business so as to attract more customers.”

“The higher the number of customers a business has, the higher the costs of the business, costs which must be invested in the maintenance and improvement of the business if it is to serve that higher number of customers.”

What matters (of course) is how costs and revenue increase as scale increases
Of course “more is better” but the question is whether cost of producing incremental quality decreases with scale.

Example: standard errors go down as the square root of sample size, a special case of diminishing returns. Twice as much data gives you 40% better accuracy.

Is this true of machine learning? Let’s see..
Figure 1. Learning Curves for Confusion Set Disambiguation

Banko and Brill, “Scaling to Very Very Large Corpora for Natural Language Disambiguation”, Microsoft Research
Voting among classifiers

Figure 3. Voting Among Classifiers

“Beyond 1 million words, little is gained by voting, and indeed on the largest training sets voting actually hurts accuracy” Banko and Brill
“... a real-case scenario of an algorithm in production at Netflix. In this case, adding more than 2 million training examples has very little to no effect.”

Learning curves for naive Bayes

\[ 2^{10} = 1,024 \]
\[ 2^{12} = 4,096 \]
\[ 2^{14} = 16,384 \]
\[ 2^{20} = 1,048,576 \]

Junqué de Fortuny Enric, Martens David, and Provost Foster. Predictive Modeling With Big Data: *Is Bigger Really Better?*, Big Data, Dec 2013, Figure 2.
“As Figure 2 [previous slide] shows, for most of the datasets the performance keeps improving even when we sample more than millions of individuals for training the models. One should note, however, that the curves do seem to show some diminishing returns to scale.”

\[2^{10} = 1,024\]
\[2^{12} = 4,096\]
\[2^{14} = 16,384\]
\[2^{20} = 1,048,576\]

Data Threshold

Solution Quality

low  threshold  high

Amount of Data

Internet-scale Data Analysis, Peter Norvig 2010
Data Threshold

Internet-scale Data Analysis, Peter Norvig 2010
Microsoft’s lament

“If Bing were bigger, it would be better…”

But if Bing were better, it would be bigger.

How to get bigger?

Imitation is the sincerest form of strategy
But is “Me too” a strategy?
Bing or Google?

Rome - Best Travel Tips on TripAdvisor - Tourism for Rome

www.tripadvisor.com - Europe - Italy - Lazio
Rome Tourism: TripAdvisor has 1,721,620 reviews of Rome Hotels, Attractions, and Restaurants making it your best Rome resource.
Hotels - Things to Do in Rome - 3 Days in Rome - Flights to Rome - Rome Travel Forum

Rome - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/ROME
Rome is a city and special comune (named "Roma Capitale") in Italy. Rome is the capital of Italy and region of Lazio. With 2.9 million residents in 1,285 km² (496.1...
Etymology - History - Government - Geography - Climate - Demographics

Videos of rome
bing.com/videos

Rick Steves' Rome: Eternally
YouTube
4:57 HD

Rome: The Rise and Fall of an Empire
YouTube
43:22

The Seven Wonders of Rome
YouTube
48:33

Rome - Do You Like This
YouTube
4:10

See more videos of rome

Rome (TV Series 2005–2007) - IMDb
www.imdb.com/title/tt0384766
★★★★★ Rating: 8.9/10 · 93,831 ratings · Action/Drama/History · 90 min
With Kevin McKidd, Ray Stevenson, Polly Walker, Kerry Condon. A down-to-earth account of the lives of both illustrious and ordinary Romans set in the last days of the...

Rome.info > Rome tourist information
www.rome.info
Offers tourist information on Rome and the Vatican city, including shopping, hotels, museums, transportation, and points of interest.

Trending on social networks about rome
twitter.com
Jim Rome on Twitter
twitter.com
The set for Shoutout. See it in action Wednesday at 8p on @SHOCyber

Rome
City
Rome is a city and special comune in Italy. Rome is the capital of Italy and region of Lazio. With 2.9 million residents in 1,285 km², it is also the country's largest and most populated comune and fourth-most populous city in the European Union by population within city limits. The Metropolitan City of Rome has a population of 4.3 million residents...

Local time: 8:36 PM 6/8/2015
Population: 2.65 million (2013)
Area: 496.26 sq miles (1,285 km²)
Travel tip: It's nicknamed the Eternal City for a reason. In Rome, +
Nearby airports: Leonardo da Vinci–Fiumicino Airport · Ciampino–G. B. Pastine International Airport
Host of: 1960 Summer Olympics
Colleges and universities: Sapienza University of Rome · John Cabot University · Pontifical North American College +

Weather
73 °F Cloudy
Bing or Google?

Rome - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Rome
Rome (/rōm/, Italian: Roma [ˈroːma] (listen), Latin: Rōma) is a city and special comune (named "Roma Capitale") in Italy. Rome is the capital of Italy and ...
Ancient Rome - Rome (disambiguation) - Rome (TV series) - History of Rome

Rome (TV Series 2005–2007) - IMDb
www.imdb.com/title/tt0384766/
Internet Movie Database
Rating: 8.9/10 - 93,879 votes
Videos. Rome -- (Series Finale) In the aftermath of the battle of Actium, Rome -- Rome is facing a dire shortage of grain, forcing Octavian to barter ...

Rome Highlights - Lonely Planet
www.lonelyplanet.com/italy/rome
Lonely Planet
A trip to Rome is as much about lapping up the lifestyle as it is gorging on art and historic sights. And there's no better way of getting into the local spirit of things ...

In the news
Lion Elevator Restored at Rome's Gladiator Colosseum
NBCNews.com - 11 hours ago
ROME, Italy - An ancient elevator that carried lions and other animals into Rome's ...

Deadly animal machine rises again at Rome's Colosseum
CNET - 54 mins ago

Dazzling jewels from an Ethiopian grave reveal 2000-year-old link to Rome
The Guardian - 2 days ago

Rome
Capital of Italy
Rome, Italy's capital, is a sprawling, cosmopolitan city with nearly 3,000 years of globally influential art, architecture and culture on display. Ancient ruins such as the Roman Forum and the Colosseum evoke the power of the former Roman Empire. Vatican City, headquarters of the Roman Catholic Church, boasts St. Peter's Basilica and the Vatican Museums, which house masterpieces such as Michelangelo's Sistine Chapel frescoes.

Founded: April 21, 753 BC
Weather: 72°F (22°C), Wind SE at 10 mph (16 km/h), 57% Humidity
Local time: Monday 8:36 PM
Hotels: 3-star averaging $130, 5-star averaging $380.
View hotels
Province: Province of Rome
Population: 2.627 million (2012) UNdata

Bing started as beta in 2010 in Germany
Came out of beta in January, 2012

Google first offered German version in 2000.
Eric Schmidt’s 40-language initiative was created in 2007

As more and more users, advertisers, and partners interact with Google across the world, the need for local products has become even more obvious. In 2007, we undertook a company-wide initiative to increase the availability of our products in multiple languages. We picked the 40 languages read by over 98% of Internet users and got going, relying heavily on open source libraries such as ICU and other internationalization technologies to design products.
Bing handles about half as many queries as Google in the US. Implications...

- experiments run at 2% rather than 1%
- experiment run for 2 days rather than 1 day
- amount of easily accessible past data is 4 weeks rather than 2 weeks

Is there some magic threshold?
Distinct queries never seen before

- "50% of queries are seen by Bing fewer than 100 times in a month" Same is true of Google.
- The fraction of queries never seen before: Nov 2008: 16%  Nov 2014: 15%
- Distinct queries: Nov 2008: 38%  Hit asymptote in 2005

“Yahoo executive vice president Hilary Schneider showed how few ads Yahoo returned for a search on [red roses in birmingham alabama]. In contrast, Google’s search results page was loaded with ads. “

There were 10 ads on Google. But 9 of the Google advertisers were also on Yahoo---their ads just weren’t showing! Several of the advertisers had the exact same ads on both search engines..

“... if Yahoo can’t target an ad for "roses" to "red roses in birmingham alabama," it’s got serious issues.”

How does Bing do with this query in 2015?
red roses in birmingham alabama

Avas Birmingham Florist - AvasFlowers.com
www.avaxflowers.com/Birmingham  (877) 804-0418
4.8 ★★★★★ rating for avasflowers.com

Best Local Flowers Delivered Save 45% Today & Get Free Delivery!
Same Day Delivery · Satisfaction Guarantee · 45% Off All Items

Get Well Flowers  Sympathy Flowers
Birthday Flowers  45% Off Sale

Images for red roses in birmingham alabama

Red Roses Delivery
www.fromyouflowers.com/Roses
4.3 ★★★★★ advertiser rating
$29.99 Stunning Rose Bouquets
“Best Value Florist” - CBS News

Red Roses from $29.99
www.proflowers.com/
4.2 ★★★★★ rating for proflowers.com
Order Red Rose Arrangements & Have Them Delivered Fresh Right Now!

Red Roses from Just $0.64
www.globalrose.com/Red-Roses
4.6 ★★★★★ rating for globalrose.com
(877) 701-7673
The Highest Flower Quality Online.
Order Now for Free Delivery!
Roses from HOOVER FLORIST - your local Birmingham, AL
... www.birminghamalflorist.com/roses.php
... your local Birmingham, AL florist and flower shop to save time and money. ... This vase of brilliant red roses is an elegant and natural way to say, ...

Red Roses in Alabama, Birmingham - local.globalrose.com
local.globalrose.com/Alabama/Birmingham/Red_Roses
Toll Free: (877) 701-7673 Hablamos Espanol; Sign In; My Account; Flower Videos; Contact Us

A Dozen Premium Red Roses in Birmingham AL - Martin ...
www.martinflowers.net/birmingham-flowers/a-dozen-premium-red-roses...
Order A Dozen Premium Red Roses from Martin Flowers, your local Birmingham florist. For fresh and fast flower delivery throughout Birmingham, AL area.

Dance with Me Bouquet with Red Roses in Birmingham AL
...
www.martinflowers.net/birmingham-flowers/dance-with-me-bouquet...
Order Dance with Me Bouquet with Red Roses dance with me from Martin Flowers, your local Birmingham florist. For fresh and fast flower delivery throughout Birmingham ...

Related searches
- Birmingham Alabama
- Club Red Birmingham AL
- Alabama Rose Song
- Salon Red Birmingham AL
- Alabama Rose Gardens
- University of Alabama at Birmingham
- Rose of Alabama YouTube
- University of Alabama Roses
Avas Avis, PA Florist - Avas Flowers
Ad - AvasFlowers.com/Avis - Avas Flowers
Best Local Flowers Delivered. Save 45% Today & Get Free Delivery!
avasflowers.com has been visited by 100K+ users in the past month
Avis Avis, PA Florist Delivery. Avas Flowers Shop
avasflowers.com is rated ★★★★★ on Bing (11,952 reviews)

Roses
Wow Them With Unique Rainbow Roses
Or Our Beautiful Assortment of Reds
Great Deals
Premium Flowers at Lowest Prices Bouquets Starting at $29.99

Sympathy Flowers
Send Your Condolences With One Of Our Many Comforting Arrangements
Admin Professional Gifts
Show Your Appreciation For Your Admin Professional The Week Of 4/20

Flower Delivery Services | Send Flowers Online …
www.avasflowers.net
Sending flowers online is easy at Avas Flowers. Our reliable flower delivery services will get your flower arrangements there fast and cheap.

Avas Flowers Complaints - 275 Consumer Reviews …
www.complaintlist.com/florists/avas-flowers
★★★★★ Rating: 1.1/5 - 275 reviews - 877-636-3033
Avas Flowers has 275 reviews or customer complaints filed and 198 comments. Consumer have reported losses from Avas Flowers of $20,458.30. Corporate Name: Avas Flowers ...

Avas Flowers Shop | Avas Flowers Shop
www.avasflowershop.com

Flowers: Blooming Love Premium Romance $99.99
1800Flowers.com

Flowers - Florist Designed Bouquet $23.99
From You Flowers

Avas Flowers Thank You Bouquet $39.99 $64.95
1-800-FLORALS

Local Alva Flowers
www.FlowerShopping.com/Alva
Same Day Alva Flower Delivery. Family Owned and Operated!
flowershopping.com is rated ★★★★★ on Bing (4,335 reviews)
Long tail queries [appeared once in a particular day]

Often misspelled, long, and local ...

country rpm bucket  query
US  recover information from digital camera hard drive
US  alpharetta air duct cleaning
US  stone mountain carpet cleaning
US  plano foundation repair
US  cheap but good auto insurance in los angels ca
US  boiler rental richmond va
US  25 mailbox hosted exchange plan
US  marietta carpet cleaner
US  new york wedding band
US  criminal attorney fort myers
US  www.mesothelioma.com
US  hosted exchange pricing
US  house tinting houston
US  disk data recovery cleveland
US  personal injury attorney, dallas
US  flower mound foundation repair
US  murder lawyer in los angeles
US  drug rehab phoenix arizona
....
So why is Google better?

1. Learning by doing is powerful

2. Search is core to Google’s business, we’ve been doing it since 1998 and have learned a lot

3. (Microsoft has learned a lot about office software, gaming, PC OSes, etc.)

4. But...
   a. Microsoft has $80B in cash to invest in search
   b. Baidu (world’s 2nd biggest search engine) is investing heavily in machine learning and foreign expansion
   c. If Google stopped innovating, we would likely see deterioration in user satisfaction within a few months
1. Google invests much more than Bing on search and ads
2. “Ballmer is willing to invest 5-10% of operating income in Web search” = $1-2B per year. [seattlepi.com, 2009.]
   a. Google’s costs are are $48B/year. Perhaps 30% of this goes to search and ads. If so, Google is spending ~10 times as much as MSFT.
   b. Some of these costs are due to larger size, of course
3. Microsoft has shown it is possible to build a me-too search engine on the cheap.
4. But how does Microsoft expect to succeed with a me-too strategy?
5. Google faces this problem in office software, but we offer a differentiated product.
Google has established a common data standard for PLAs.

"I want my ads to grab a customer’s attention."

A Product Ad takes up more space on the search results page than a traditional text ad. Product Ads can also show simultaneously with your text ads, giving you even more real estate.

"I’m busy — I need to free up time so I can focus on my business. Efficiency and scalability are key."

Product Ads can save you time. Product Ads pull information from your product feed to generate your ad — like image, price and brand name.

"I need to be efficient — I’m already advertising with Google Product Listing Ads."

If you are already running Google Product Listing Ads, it’s easy to import those campaigns into your Bing Ads account. Once set up, Product Ads are easy to manage and maintain.
More Volume for Product Ads in Time for Your Holiday SEM Campaigns

Product Ads now served on 100% of eligible Yahoo Bing Network search traffic. Product Ads on the Yahoo Bing Network have been available since March, and search advertisers have been seeing great results.

In case you’re not using them yet, Product Ads appear for searches that are relevant to your product inventory and automatically serve the appropriate images, text, and pricing from your product catalog. Advertisers have been asking for one key enhancement---more volume, so they could scale those great results. We’re happy to announce that Product Ads are now showing on 100% of eligible Yahoo search traffic, giving advertisers a boost just in time for the holidays.
Google Confidential and Proprietary
Microsoft VP Rik van der Kooi:

“We spend quite a bit of time thinking about the operating system level signals that we have access to. We do not use them in Bing. ... But theoretically we have access to all of the signals of what a user does on the operating system and on the computer, irrespective of what browser they use, etc. There is a richness of additional information there that we don’t leverage today.”
Microsoft Corporation
NASDAQ: MSFT - Jun 24 7:59 PM EDT

45.63  + 0.28  (0.60%)  
After-hours: 45.73  + 0.10  (0.21%)  

1 day  5 day  1 month  3 months  1 year  5 years  max

Open  45.67  High  46.24  Low  45.55  
Market cap  371.39B  P/E ratio (ttm) 18.96  Dividend yield  2.72%

MSFT: Summary for Microsoft Corporation- Yahoo! Finance
finance.yahoo.com/q?s=MSFT
View the basic MSFT stock chart on Yahoo! Finance. Change the date range, chart type and compare Microsoft Corporation against other companies.
Yahoo! Finance MSFT chart - Historical Prices - MSFT Key Statistics - Major Holders

Microsoft Corporation: NASDAQ:MSFT quotes & news ...
www.google.com/finance?cid=358464
Get detailed financial information on Microsoft Corporation (NASDAQ:MSFT) including real-time stock quotes, historical charts & financial news, all for free!