Scaling the API ata Fo conom Selec Document rmedia Pa Nine Esse and your H cumenting edia Past a ne Essentia n Controls ting the Pro our Hypern vour Hvp menting yo g the Prope our Hypern E**ssential H** vour Hype a Past and ontrols to v Hypermed With Scale-Free Networks ast and Hyperme per Data Fo o vour Res ng Protocopermedia ır Hyperm our Respoi Data For ProtocolN Hypermed Mike Amundsen **Principal API Architect** Layer 7 Technologies @mamund @mamund #ScaleFreeAPIs

I come as a messenger...





The Goddess Iris

Iris is associated with

 communication
 new endeavors





I bring hard news...





Iridium (lr)

- Very hard & brittle
- Second-densest element
- Most corrosion-resistant
- Often found within impact craters
- Used in deep space satellites







 The most common approach to Internet-based APIs runs counter to a fundamental property the Web

f="bttp://www.exuppet"

subm



title="company logo"

 At our current pace of growth, we may not be able to keep this up for much longer.



title="company logo"

• We may be headed for:

"Scale-pocalypse!"





title="company logo"

#ScaleFreeAPIs

But wait,

I'm getting ahead of myself...





edia Past a ia Past and ur Hyperm **Past and Pre** media App Proper Date col Method s that Unde vour Hype t and Prese per Data F dia Applica **Methods** t Hypermedi ur Hyperm nd Present rmedia Fac r Data Forn permedia Applicatio ethods to v Hypermed ata Format ods to you Hypermed Serving Ur permedia Selecting t ication Coi rving Up yc to your H Document rmedia Pas Nine Essei tion Contro your Hype ng Up your and your H edia Past a ne Essentia ting the Pro n Controls our Hypern Up your <u>Hy</u> your Hype menting yo ia Past and Essential H your Hype g the Prope ontrols to ^v Hypermed our Hypern **Binding Puur Hyperm** ast and Preential Hype trols to you ng your Hy Hypermed t and Prese tial Hypern nding ProteHypermed Proper Data ls to your F permedia ng Protoco permedia nd Present per Data Fe o vour Res ur Hyperm Data Scale-pocalypse 1998 Hypermed permedia) rmedia Fac atal rmedia Pas edia Past a a Past and Nine Esser @mamund #ScaleFreeAPIs

1973 – Lansing, Michigan

- I start my first year at Michigan State University
 - The same year, same city, Larry Page is born.





1994 – Stanford University

• Jerry Yang & David Filo

"Jerry's guide to the world wide web"
YAHOO!



tle="company logo"

Curated, hierarchical

index



1998 – Stanford University

• Sergey Brin & Larry Page

The Anatomy of a Large-Scale Hypertextual Web Search Engine (1998)

"...due to rapid advance in technology and web proliferation, creating a web search engine today is very different from three years ago."

//www.exupdated>

C_12th 13p18;20:02z</updated

The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Computer Science Department, Stanford University, Stanford, CA 94305, USA sergey@cs.stanford.edu and page@cs.stanford.edu

Abstrac

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/ To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on the web, very little academic research has been done on them. Furthermore, due to rapid advance in logy and web proliferation, creating a web search engine today is very different from three years ago. This paper provides an in-depth description of our large-scale web search engine -- the first such detailed public description we know of to date. Apart from the problems of scaling traditional search techniques to data of this magnitude, there are new technical challenges involved with using the additional information present in hypertext to produce better search results. This paper addresses this question of how to build a practical large-scale system which can exploit the additional information present in hypertext. Also we look at the problem of how to effectively deal with uncontrolled hypertext collections where anyone can publish anything they want

Keywords

title≃"company logo"

World Wide Web, Search Engines, Information Retrieval, PageRank, Google

1. Introduction

(Note: There are two versions of this paper -- a longer full version and a shorter printed version. The full version is available on the web and the conference CD-ROM.)

The web creates new challenges for information renieval. The amount of information on the web is growing spajibly, as well as the number of new users insexperienced in the at of web research. People are likely to surf the web using its link graph, often starting with high quality human maintained indices such as Yahool or with search engines. Human maintained lists cover popular topics effectively but are subjective, expensive to build and maintain, slow to improve, and cannot cover all scotteric topics. Automated search engines that rely on keyword matching usually return too many low quality matches. To make matters wores, some advertisers attempt to gain people's attention by taking measures meant to mislead automated search engines. We have built a large-scale search engine which addresses many of the problems of existing systems. It make especially heavy use of the additional structure present in hypertext to provide much higher quality search result. We chose our system name, Google, because it is a common speling of google, or 10¹⁰⁰ and fits well with our goal of building very large-scale search

What was different?

What did Brin & Page perceive that others had missed?
How did Google avoid the "Scale-pocalypse"?



ential Hyp tial Hypern ne Proper I col Methodur Hyperm's that Unde Hypermed Proper Dat media App vour Hype and Prese vpermedia **Methods** tHypermedi per Data F dia Applica ur Hyperm rmedia Facr Data Forn ethods to vpermedia a Applicatio Understan Hypermed edia Factorata Format ods to you Serving Up oplication permedia permedia ia Factors ication Cor Selecting t rving Up yc to your H Document rmedia Pas Nine Esser tion Contro lecting the your Hype ng Up your and your H edia Past a ne Essentia ting the Pro n Controls our Hypern Up your <u>Hy</u> menting yo vour Hype Essential H g the Prope our Hypern ia Past and ontrols to your Hype Hypermed nting your ential Hypere Proper I **Past and Pr** trols to you **Binding Prur Hyperm** Hypermed ng your Hy and Presetial HypernProper Data nding ProteHypermedi ls to vour F permedia nd Present Hypermed per Data Feo your Res ng Protocopermedia ur Hyperm in "Scale-Free permedia – Data Fornygır Respo ree as rmedia Fa edia Facto col Method ia Factors Nine Esser @mamund #ScaleFreeAPIs

Scale-Free Network

• "A network whose degree distribution follows a power law." $P(k) \sim k^{-\gamma}$ f="http://www.examplate



title≃"company logo"

bidat fon and

Scale-Free Network

• "A network whose degree distribution follows a *power* law." $P(k) \sim k^{-\gamma}$ f="http://www.examplate bidat fon and



title≃"company logo"

Power Law



"Long Tail"



"Long Tail"

- Not an equal distribution
- A few nodes have many of links
 Many nodes have a few links
 "Preferential Attachment" – Barabási–Albert (1999)



Power Law Web Search

Brin & Page recognized that
curation and aggregation
would not scale on the WWW



title="company logo"

Power Law Web Search

• Instead, they decided to allow users and the links they create to inform the search engine.



Power Law Success

• Using this feature of the WWW allowed Google to gain the leading role in search • and build a multibillion dollar company.



ential Hypig the Prope ne Proper I col Methodur Hyperm Proper Datamedia App s that Unde vour Hype per Data Fedia Applica and Prese Methods t Hypermedi ur Hyperm rmedia Fac<mark>r Data Forna Applicatio</mark> ethods to vpermedia Understan Hypermed edia Factor ods to you ata Format pplication Serving Up Hypermed permedia. **Selecting trication Con** rving Up yo to your H Document rmedia Pas lecting the tion Contro your Hype ng Up your and your H ne Essentia our Hypern edia Past a ting the Pren Controls Up your <u>Hy</u> your Hype menting yo Essential H g the Prope ontrols to **y** our Hypern your Hype ia Past and Hypermed Past and Pro ential Hyp<mark>c</mark>ie Proper I trols to you **Binding Puur Hyperm** Hypermed ng your <u>Hy</u> and Prese tial HypernProper Datals to your **F** nding ProteHypermedi permedia ng Protocopermedia Hypermed per Data Feo your Res ur Hyperm Tave We forgotten Serving Up iting Clien Hyper? Hyperro T e Selecting lecting the ting the Pr g the Prope @mamund #ScaleFreeAPIs

Hub-Vulnerability

- Unexpected disasters can cause complex systems to fail.
- DNS failures of the 2000s
- Recent Amazon AWS failures
- 2008 Monetary system failures

Too big to fail"



title="company logo"

Node-Resiliency

• Highly distributed systems reduce risk of system-wide failures.



itle≃"company logo"

Does power-law only apply to search?







Collaboration



Identity





APIs?

Rordstforap

type="text" <label>Ke value=""

 Do we really want to add more vulnerability to the WWW?

Can we afford to rely on curation and aggregation for APIs?

f="bttp://www.exampdated



title="company logo" />



What if we changed our model?

- Could we create nodebased services?
 - What would we gain?
 - What would we lose?



Business models based on nodes

- Powerful client aggregators
 User-centric – Discovery
 - Linking

Sharing



Allow users to identify

new, popular services



Advantages of node-based models

- Increased processing power (SETI)
- Access to more metadata (EXIF)
- User-driven selection models (reverse-flow)
- Reverse scale/vulnerability challenge
- Increase "network" intelligence
 Mix of "smart" and "dumb"
 - clients (IoT)



Yes, but can you make money this way?





Models/Paradigms

- IBM gladly let Microsoft own the OS (*IBM knew* hardware was the real profit source)
 - Yahoo knew the value of curated indexes (*Google had another idea*)
 - Few thought there was a business in Open Source movement.



ProtocolM otocol Metl col Methocur Hyperm's that Unde **Proper Dat** media App Hypermed vour Hype l Methods t and Prese per Data F dia Applica Hypermedi nat Underst ur Hyperm rmedia Faor Data Forr ethods to y permedia a Applicatic Understan Hypermed edia Factor oplication (ods to you Serving Up ata Format Selecting t ication Cor rving Up yo s to your H Document rmedia Pas tion Contro your Hypeng Up your and your H edia Past a ne Essentia ting the Pro n Controls our Hypern Up your Hy your Hype menting yo g the Prope Essential H Hypermed your Hype ia Past and ontrols to v our Hypern Past and Property of the Prope ential Hype le Proper I trols to you **Binding Prur Hyperm** ng your Hy Hypermed tial Hypern and Prese Proper Data ls to your F nding ProteHypermedi permedia ng Protoco permedia ur Hyperm nd Present per Data Feo your Res Be a Node and not a Hub! permedia col Method Methods t ethods to y @mamund #ScaleFreeAPIs

Node-based?

didat for aber

'text" <label> value='

- Twitter
- FacebookGoogle DocsWordpress

f="bttp://www.example



Node-based?

• I post messages to my own machine The client sends it to one or more selected services of my choice • Publishers compete for my content Publishers distribute to other nodes



Avoiding the Scale-pocalypse

- API providers:
- Consider hub vulnerability problem
 - Explore node-based advantages

Empower users and reap rewards



company logo

Avoiding the Scale-pocalypse

- API consumers:
- Is there a node-based provider available?
 Build client-based aggregators
 Empower users and reap rewards



Protocol MUp your HyClients that nting your your Hypeents that Uping your Hy col Methodur Hyperms that Undeyour Hype media App Hypermed Proper Data Hypermedinat Underst and Prese per Data F dia Applica l Methods t ur Hyperm rmedia Faor Data Forr a Applicatio ethods to y permedia Hypermed ata Format Serving Up oplication ods to you derstand v Selecting t rving Up yc s to your H your Hypeng Up your rmedia Pas tion Contro and your H edia Past a ne Essentia ting the Pro n Controls our Hypern Up your Hy l your Hype menting yo your Hype ia Past and Essential H our Hypern g the Prope ontrols to Hypermed **Binding Prur Hyperm** Hypermed Past and Pre ential Hyp trols to you ng your Hy and Prese tial Hypern nding ProteHypermedi Proper Data ls to your F permedia ur Hyperm per Data Fe ng Protoco permedia o your Res our Respo Serving Upriting Clien Hypermed rving UA ve rmedia Fac otocol Metl nds **9** rmedia ata Format col Methoong Up your Clients that Document **Up your Hy** ecting the cumenting your Hype @mamund #ScaleFreeAPIs

So, my message is a positive one...

• We have an opportunity for new endeavors







So, my message is a positive one...

 We can build strong, vulnerability-resistant systems



title="company logo"

The news is not so bad...





The news is not so bad...

• There is a way to build systems that take advantage of a key principal of the WWW



So, the news is no so bad...

 Using power-law and node-based systems, we can continue to meet growing demand.



The news is not so bad...

bidat fon aber

• And we can avoid

"Scale-pocalypse!"

f="http://www.exappa



title="company logo"

Be a Node!





Scaling the A at Unders ·Data Form Applicatic ethods to y permedia Hypermed Serving U Economy ata Form a Factors Document rmedia Pas tion Contro vour Hype ng Up vour and your H cumenting edia Past a ne Essentia ting the Pro n Controls our Hypern Jp vour H[·] **vour** Hype menting yo g the Prope E**ssential H** a Past and vour Hype ontrols to v Hypermed ur Hypern ast and Pr ential Hyd trols to you Binding Prur Hypern <u>ne Proper I</u> ng your Hy Hypermed With Scale-Free Netw and Prese ng Protoco **per Data F**e o vour Res permedia our Respoi Protocol M Serving U ata Format Mike Amundsen **Principal API Architect** Layer 7 Technologies @mamund @mamund #ScaleFreeAPIs

References

- http://en.wikipedia.org/wiki/File:Winged goddess Cdm Paris 392.jpg
- http://en.wikipedia.org/wiki/File:Iridium-2.jpg

"ethod="post"

٠

- http://dynamicdecorator.files.wordpress.com/2013/01/fesvh.png
- http://www.flickr.com/photos/luvataciousskull/3374223404/in/photolist-696AXi-69aM4A-69aM51-
- 69aM7h-69aM8o-69aM95-69aM9S-69aMaj-69aMbC-69aMdW-69aMeC-69aMgf-69aMgE-69aMjj-6c5PSD-6kLGyn-6uuq6s-6CazEo-6CDvfQ-6PtMEG-6ZBpBx-7f9XeB-7keQgx-7kzMec-7kzMmB-7kDDmN-7kDDxb-7kDDDh-7rtS3aePziXc-9eQeak-9eTm8W-9eQdCr-9eQcYk-cwkcSy-eS88UX-8pb5fQ-9D4KUU-884Y12-eRZG7k-ejqjh3-cnWzM3-19GaC-84WtXY-88WN2g-bqxc3k-9LjgRw-8GtxN8-8Aw55
- http://3.bp.blogspot.com/ VyTCyizgrHs/S-Lp0UmtOVI/AAAAAAAHik/HSQ1KfRLRa0/s1600/ericsson1.jpg
- http://www.flickr.com/photos/34206904@N02/6841249195/in/photolist-baxc3k-9LigRw-8GtxN8-8Aw55k-9G6H db
- http://kinlane-productions.s3.amazonaws.com/api-evangelist-site/blog/apis-resource-to-last-mile.png
- http://blog.adw.org/wp-content/uploads/question-mark.jpg
- http://upload.wikimedia.org/wikipedia/commons/a/ac/Jerry Yang and David Filo.jpg
- http://en.wikipedia.org/wiki/File:Larry Page in the European Parliament, 17.06.2009.jpg
- http://upload.wikimedia.org/wikipedia/commons/d/d2/Internet_map_1024.jpg
- http://upload.wikimedia.org/wikipedia/commons/8/8a/Long tail.svg
- http://i.msdn.microsoft.com/dynimg/IC76141.gif
- http://dangerouslvirrelevant.org/images/various-small/Linked thumb1.jpg
- techcium h2011.files.wordpress.com/2009/08/google.jpg
 - icflickr.com/4150/5018184807_60be8b669e_z.jpg



References

- http://www.cesa.eu/images/timeline/pictures/09_2008.jpg
- http://www.codeproject.com/KB/web-cache/AdvanceCaching/partioned.jpg
- http://blog.mybackupbox.com/wp-content/uplo

٠

٠

iethod="post"

- https://si0.twimg.com/profile_images/480262550/box.png ads/2012/06/google_drive.jpg
- http://svn.apache.org/repos/asf/subversion/trunk/notes/logo/256-colour/subversion_logo-384x332.png
- et/assets.github.com%2Fimages%2Fgravatars%2Fgravatar-user-420.png
- http://3.bp.blogspot.com/-ZeYZpE3tESU/UOz8-49WYvI/AAAAAAAAAAAAP8/vt0-vDzXxwk/s72-c/database-develope r-Eastbourne-Brighton-Sussex.png
- http://hueniverse.com/wp-content/uploads/2009/09/OAuth-Shine-300x298.png
- $\underline{http://mindcommerce.files.wordpress.com/2013/05/telephonyapi_management.gif?w=\!286\&h=\!300$
- http://blogs.bmj.com/bmj-journals-development-blog/files/2010/12/semantic-web.jpg
- http://blog.caplinked.com/wp-content/uploads/2011/09/iStock_000009230566Large.jpg
- http://studsovet.ipsa.kpi.ua/uploads/posts/2013-05/1368982043_ibm-intel-logo.jpg http://www.shinyshiny.tv/microsoft-logo 111129012732.jpg
- http://rack.2.mshcdn.com/media/ZgkyMDEyLzEyLzA0L2I1L3lhaG9vdm9pY2VzLmJrSC5qcGcKcAl0aHVtYgk5NT B4NTM0IwplCWpwZw/61883c7e/479/yahoo-voices-hacked-450-000-passwords-posted-online-7169a7e88d.

http://nyocommercialobserver.files.wordpress.com/2012/06/google-logo.gif?w=400&h=283

http://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Opensource.svg/220px-Opensource.svg.png http://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Opensource.svg/220px-Opensource.svg.png http://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Opensource.svg/220px-Opensource.svg.png http://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Opensource.svg/220px-Opensource.svg.png http://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Opensource.svg/220px-Opensource.svg.png http://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Opensource.svg/220px-Opensource.svg.png #Apy/ExRs_respect.com/images/xml-web-services-api.jpg chnologies http://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Opensource.svg/220px-Opensource.svg.Png #ScaleFreeAPIs