

UNIVERSITY OF  
Southampton

# History of Search Engines

COMP3220 Web Infrastructure

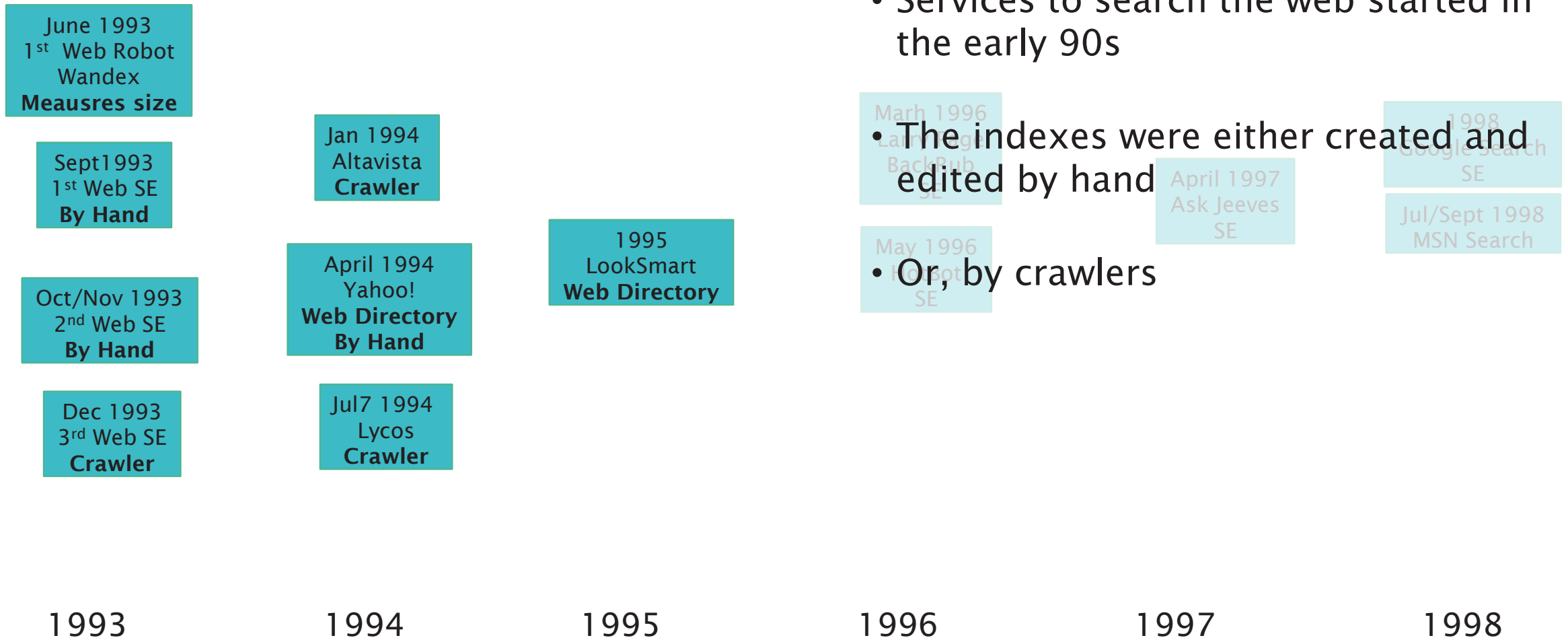
Dr Heather Packer – [hp3@ecs.soton.ac.uk](mailto:hp3@ecs.soton.ac.uk)

## Before Search Engines

- Exploit hyperlink structure
  - Personal homepages with links
  - Directories
- Word of mouth
  - Email, forums, Usenet



# Timeline 90's - Hand vs Crawler

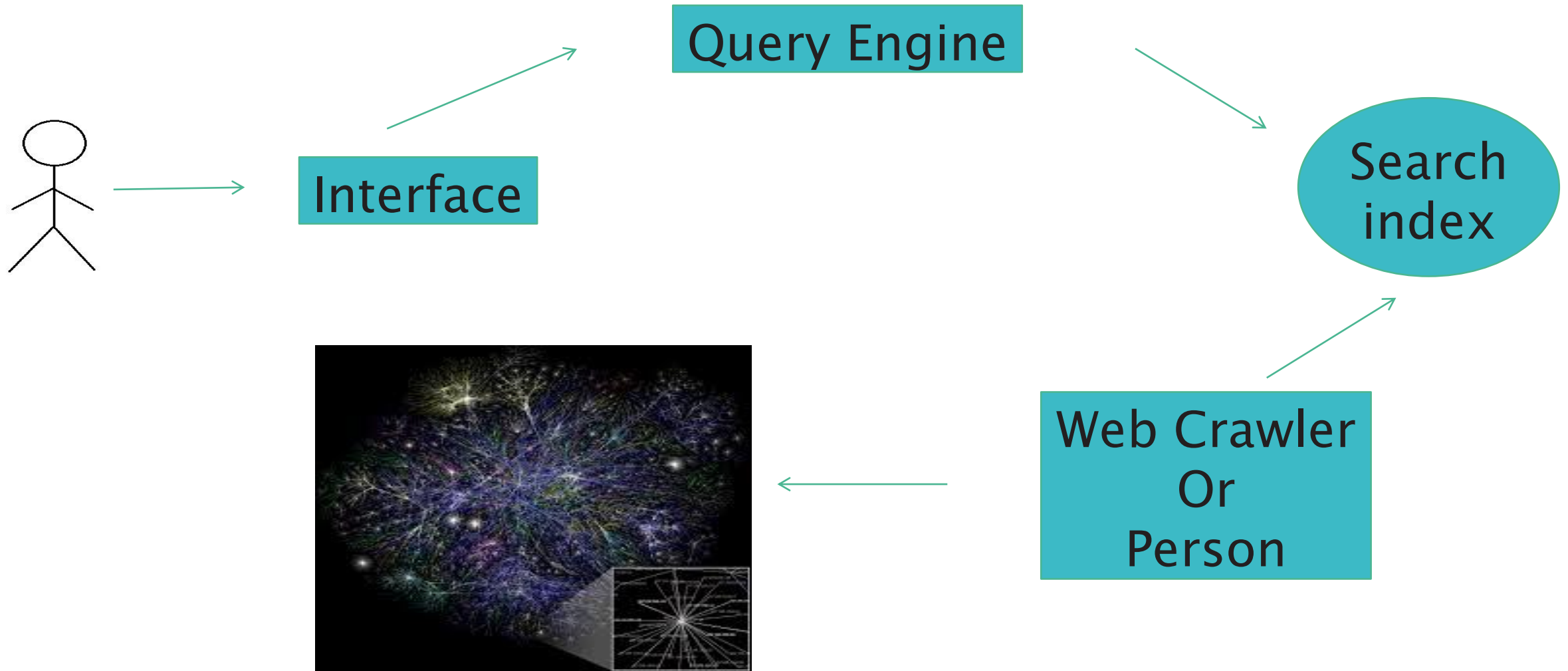


- Services to search the web started in the early 90s

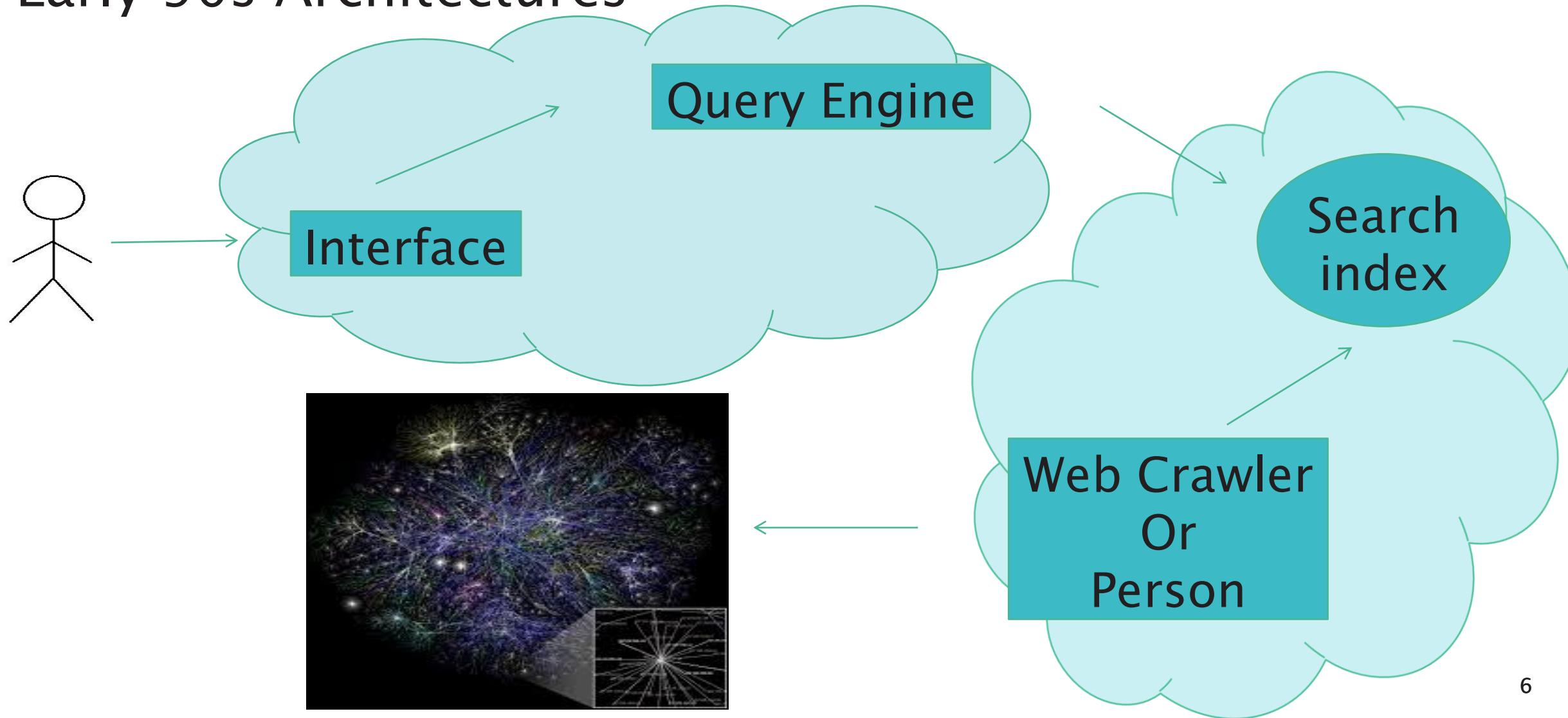
- The indexes were either created and edited by hand

- Or, by crawlers

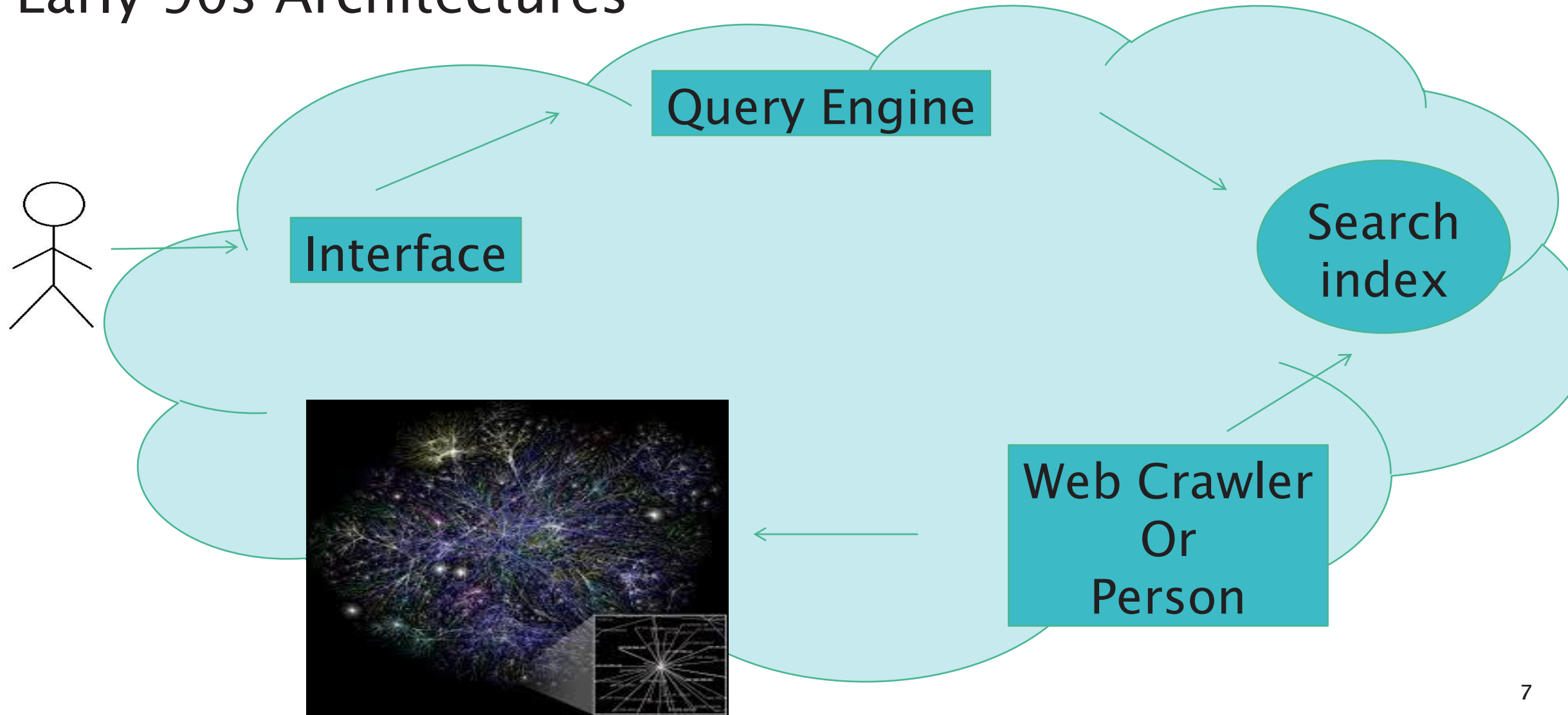
# Early 90s Architectures



# Early 90s Architectures



# Early 90s Architectures



# Web Crawlers

- Wandex 1993 – size of the web
- WebCrawler Dec 1993 - indexer
- *WebCrawler* 1994 - indexed entire web page







# WEBCRAWLER™

To search the WebCrawler database, type in your search keywords here. Type as many relevant keywords as possible; it will help to uniquely identify what you're looking for.

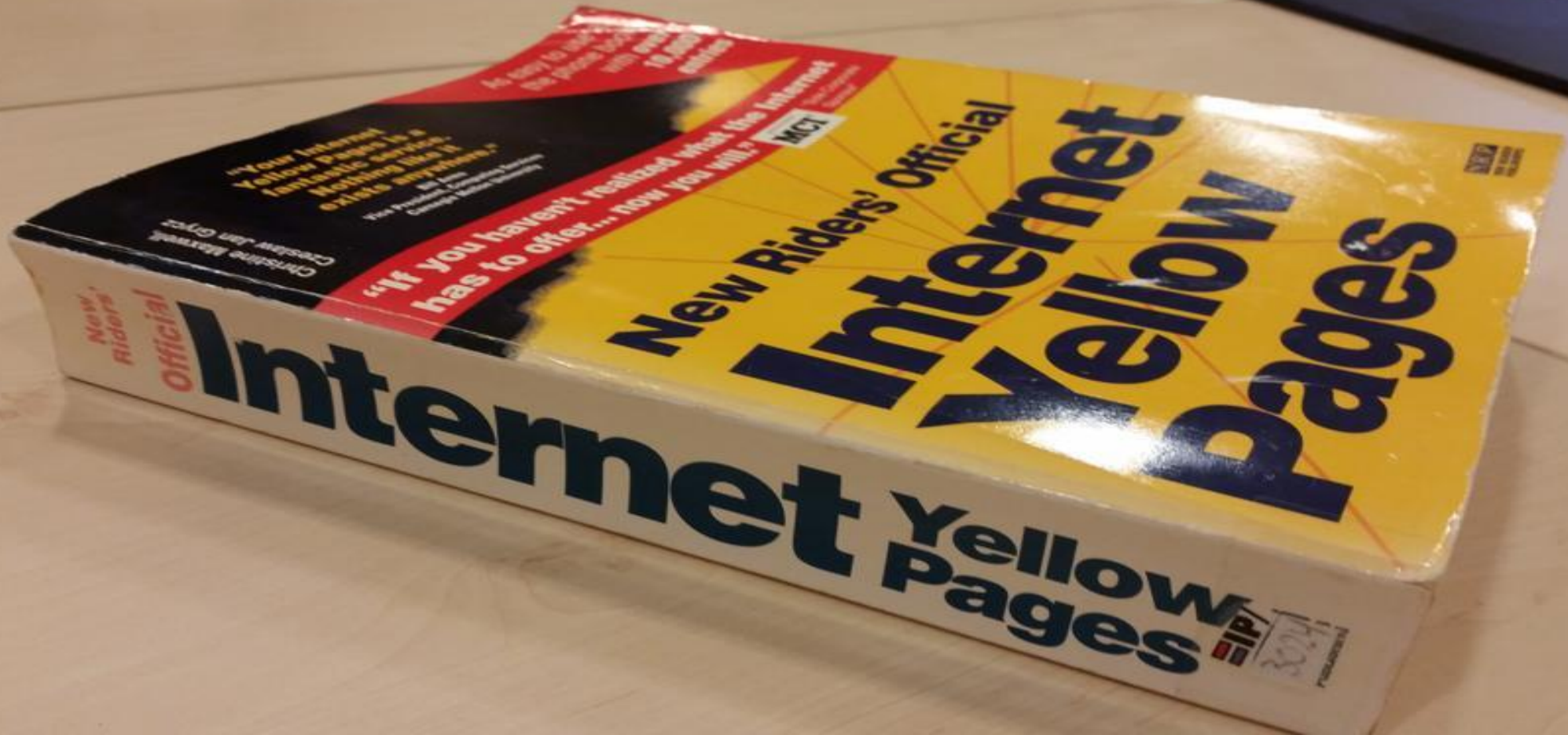
AND words together

Number of results to return:

# Yahoo!



# Internet Yellow Pages 1994





# Website - Telegraph 1994

Electronic Telegraph 15 November, 1994 THE FRONT PAGE

0 11 00 10 01 11 01 00 11 00 10 01  
electronic **Telegraph**  **The Front Page**

 **Tories block open inquiry on MPs' cash**

LABOUR failed last night to force the Commons privileges committee to hold hearings into "cash-for-questions" allegations against Tory MPs in public rather than private. The Government used its majority to block the Labour demand for an open investigation, despite warnings from some senior Conservatives that the row over standards in public life was damaging Parliament's standing.

 **Rebels threaten to behead Britons**

THREE young British tourists have been kidnapped by Kashmiri rebels who threatened to behead them if Indian authorities do not release nine Islamic militants from jail.

# Website - New York Times 1995

Electronic Telegraph 15 November, 1994 THE FRONT PAGE

9 11 00 10 01 11 02 00 11 00 12 01

electro

 To

LABOUR fa  
hearings int  
private. The  
investigatio  
standards in

 Re

THREE you  
threatened t  
from jail.

**NEW**  
**AUTO**  
**CLASSIFIEDS**

# The New York Times

ON THE WEB

knicks <sup>NEW!</sup>  
celtics  
one on one  
newyork.boston.com

"All the News That's Fit to Print" Tuesday, November 12, 1996

**SECTIONS**

- Front Page • CyberTimes
- Politics • Business
- Editorials/Letters • Op-Ed
- Arts & Leisure • Travel
- Real Estate • Job Market
- Diversions • Web Specials

**NEWS BY CATEGORY**

**CLASSIFIEDS | FORUMS**

**SERVICES | SEARCH**

**TABLE OF CONTENTS**



Crowds Looking for Food in Zaire

**LATE NEWS UPDATE**

## Two Aircraft Collide In Air Near Delhi, Killing About 350

**IN CYBERTIMES**

## Europe Betting on Self-Regulation to Control the Internet



# Website - BBC News 1997

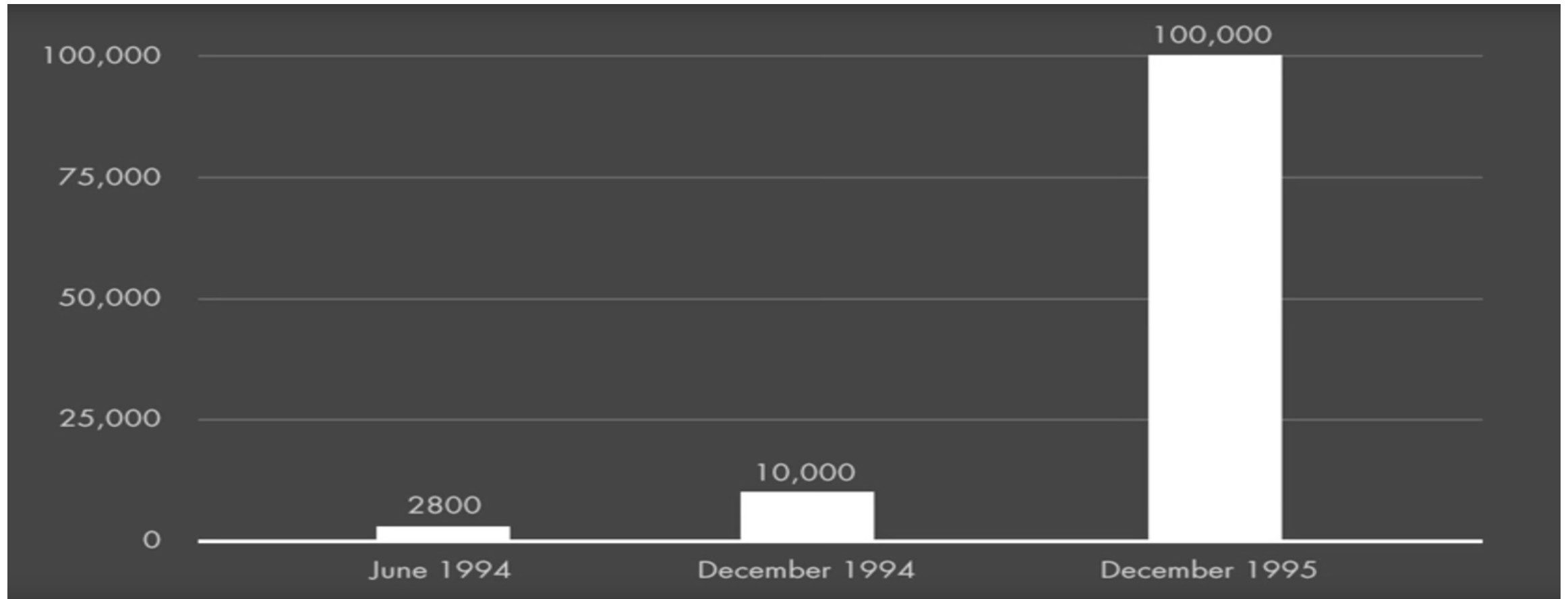
The screenshot shows the BBC News website interface from 1997. At the top, it displays the date "15 November, 1994" and "THE FRONT PAGE". The main header features the "BBC NEWS" logo with the tagline "UPDATED EVERY MINUTE OF EVERY DAY". Navigation links include "HOME PAGE", "SITEMAP", "SCHEDULES", "BBC INFORMATION", "BBC EDUCATION", and "BBC WORLD SERVICE".

The left sidebar contains a "NEW AUTO CLASSIFIEDS" banner and a list of "SECTIONS" including Front Page, Politics, Business, Arts & Leisure, Real Estate, Job, Diversions, and Web. Below this is a "NEWS BY CATEGORY" section with links for CLASSIFIEDS and SERVICES.

The main content area is titled "Front Page" and features a large article titled "Tax splits EU" with a photo of a man speaking. The article text reads: "Plans to harmonise tax policies across the European Union are placing its members' finance ministers on a collision course." Below the article are sub-headlines: "EU finance ministers split", "Tax harmony within EU?", and "Brown threatens veto over tax".

Other visible elements include a "News in Audio" and "News in Video" section, a "Newyddion" (Welsh) section, and a "Talking Point" section at the bottom. A "World Aids Day" banner is visible in the bottom right corner.

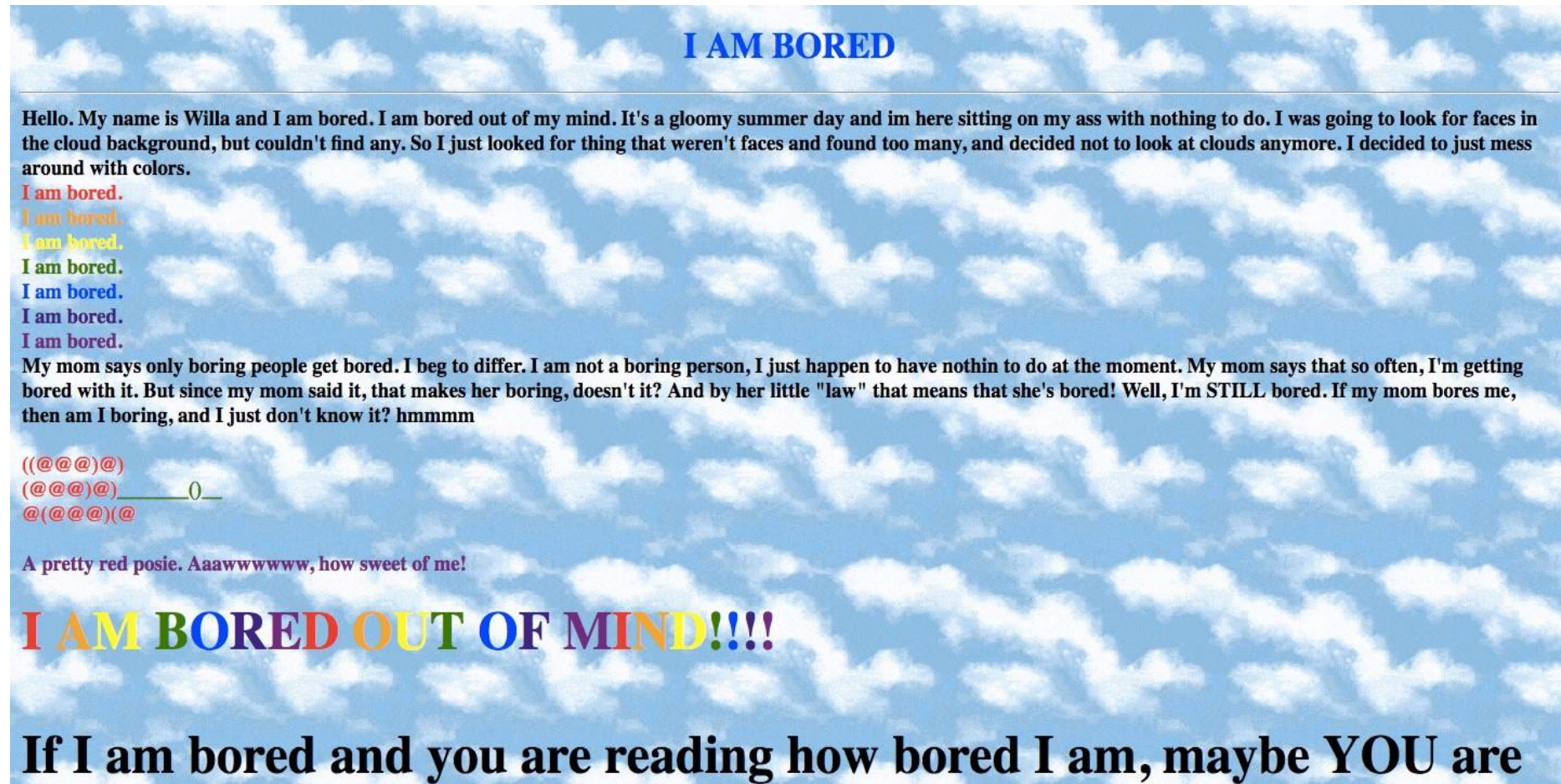
## Number of Websites June 1994 - Dec 1995





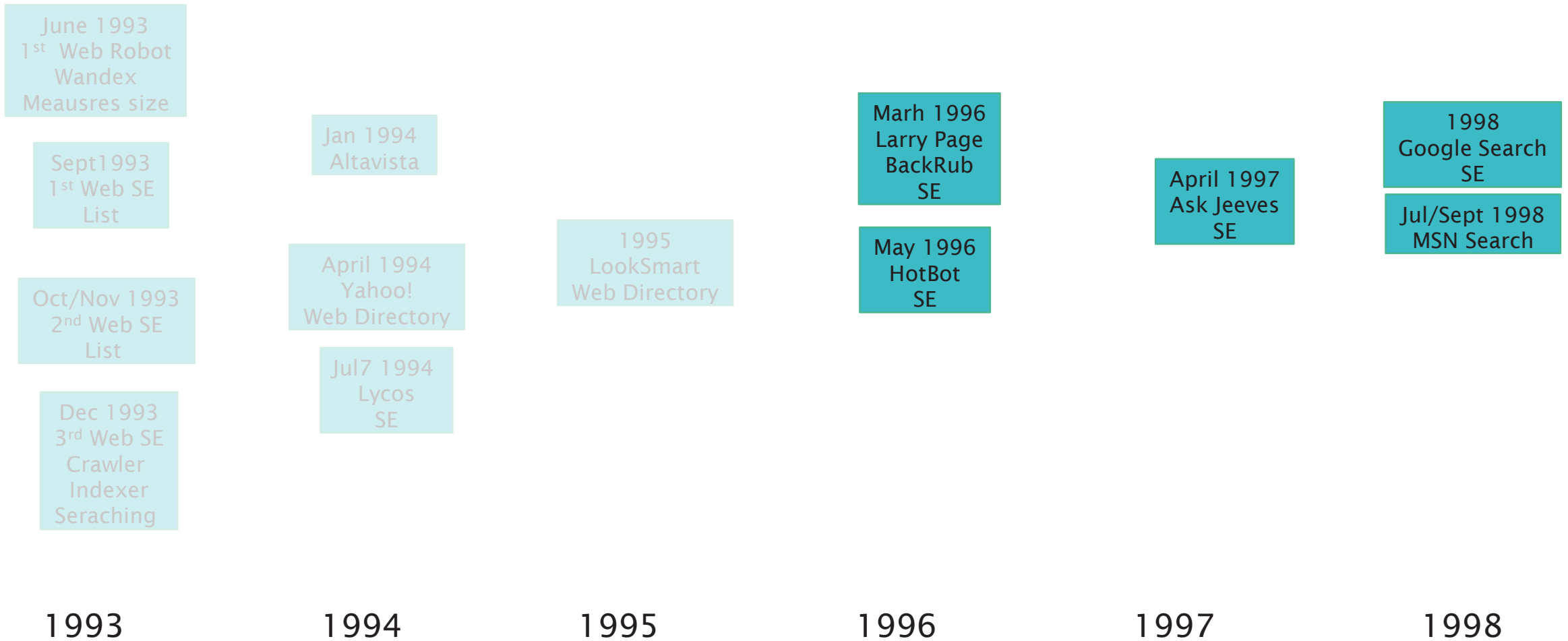
# Search Engines Issues Mid '90s

- Scale
  - How to rank pages to give the best results
- Spamming
  - Web searches favoured web pages with high keyword density
- Keywords
  - Spelling mistakes





# Timeline 90's



# BackRub

- Utilised PageRank
  - More advanced than previous indexing
- Used back links



**BackRub** is a "web crawler" which is designed to traverse the web.

Currently we are developing techniques to improve web search engines. We will make various services available as soon as possible.

**Sorry, many services are unavailable due to a local network failure beyond our control. We are working to fix the problem.**

# The Original Google Storage

- Larry Page needed a large amount of disk space to test PageRank
- 10 4GB hard drives



## The Original Google

In 1996 Larry Page and Sergey Brin, students in Stanford CSD, needed a large amount of disk space to test their PageRank™ algorithm on a large amount of wide-web data. At that time hard disks were the largest storage technology they assembled 10 of these hard disks in a low-cost cabinet.

In Nov 1999, Google Inc. by then one of the primary search engines on the web, provided replacement capacity to the Digital Library Project that we could move this original assembly into our history department.

As of September 2000, Google, located in Mountain View, operated a search engine, searching and web crawling, on a LINUX operating system.

# Google 1997-8

The Google logo is displayed in its classic multi-colored font: 'G' is blue, the first 'o' is red, the second 'o' is yellow, 'g' is blue, 'l' is green, 'e' is red, and the exclamation point is blue.

Search the web using Google!

10 results



Google Search

I'm feeling lucky

*Index contains ~25 million pages (soon to be much bigger)*

**[About Google!](#)**

[Stanford Search](#) [Linux Search](#)

Get Google! updates monthly!

your e-mail

Subscribe

[Archive](#)

Copyright ©1997-8 Stanford University



# The Evolution of Search Engine Features

- Search Engine algorithms are constantly changing
- They strive to improve their:
  - Efficiency
  - How to order the results
    - Freshness
    - Relevancy
  - How to improve user interaction
  - Additional features

# Google Adwords 2000 Funding

Google  Search [Advanced Search](#) [Preferences](#)

Web [Show options...](#) Results 1 - 10 of about 22,100,000 for **bird houses** (0.19)

**Bird Houses**  
www.Scotts.com Scotts attracts colorful birds to your backyard!

**Specialty Bird Houses**  
www.birds-out-back.com Roosting Boxes, Purple Martins, Bat Chalets.

**Bird Houses at BestNest**  
www.bestnest.com Over 225 different houses in stock. Free shipping!

[Learn More About Bird Houses](#) [↑] [×]

**Bird House:** JustBirdHouses.net learn more about blue bird houses and birdhouses, along with our blue bird house and purple martin bird houses.  
www.justbirdhouses.net/ - [Cached](#) - [Similar](#) - [🗨]

**Bird Houses** [↑] [×]  
Bird Houses and Bird Feeders for north american bird species.  
www.birdhouses101.com/ - [Cached](#) - [Similar](#) - [🗨]

**Bird Feeders, Bird Houses - The Backyard Bird Company**  
Bird Feeders - The Backyard Bird Company has a variety of bird feeders will accent your landscape and attract wildlife.  
[Bird Houses](#) - [Birdfeeders](#) - [Decorative Bird Houses](#)  
www.backyardbird.com/ - [Cached](#) - [Similar](#) - [🗨]

Sponsored Links

**Bird Houses**  
Find All Types Of Bird Feeders And Houses At Lowe's® New Lower Price  
www.Lowes.com

**Bird Houses Sale**  
Authorized Dealer - New Designs. Low Price Guarantee- Free Shipping.  
www.OutdoorLivingShowroom.com  
[Google] [Checkout]

**High Quality Bird Houses**  
Nesting boxes & decorative houses. 5-Star Service. Free Shipping \$75+  
www.backyardbird.com

**Decorative Bird Houses**  
Beautify Your Garden With Our Wooden Bird Houses at a Discount.  
BirdHouseStation.com  
[Google] [Checkout]

# Relevancy

- During 9/11 it was apparent that search engines did not cater of time-sensitive queries
- This lead to Google developing Google News
- The freshness of search engine's index became more important



# Google 2012

- In 1999, it took Google one month to crawl and build an index of about 50 million pages.\*
- In 2012, the same task was accomplished in less than one minute.\*

\*Mitchell, Jon. "How Google Search Really Works." Readwrite. February 29, 2012.



# Search Verticals

- In a bid for content and market share, Search Engines develop other search verticals:
  - Books
  - Travel
  - Finance
  - Shopping
  - Scholar

# Mobile Search

- By 2015 mobile accounted for more than half of all searches
- Searches need to be context aware (location)
- Limited screen and bandwidth means relevancy is critical
- Boost mobile-friendly pages in ranking algorithm

# Outcome

- Search engines were built out of necessity
- Search engines features where developed due to demand or world events