

I was going to talk about

“Information Retrieval in extreme contexts”

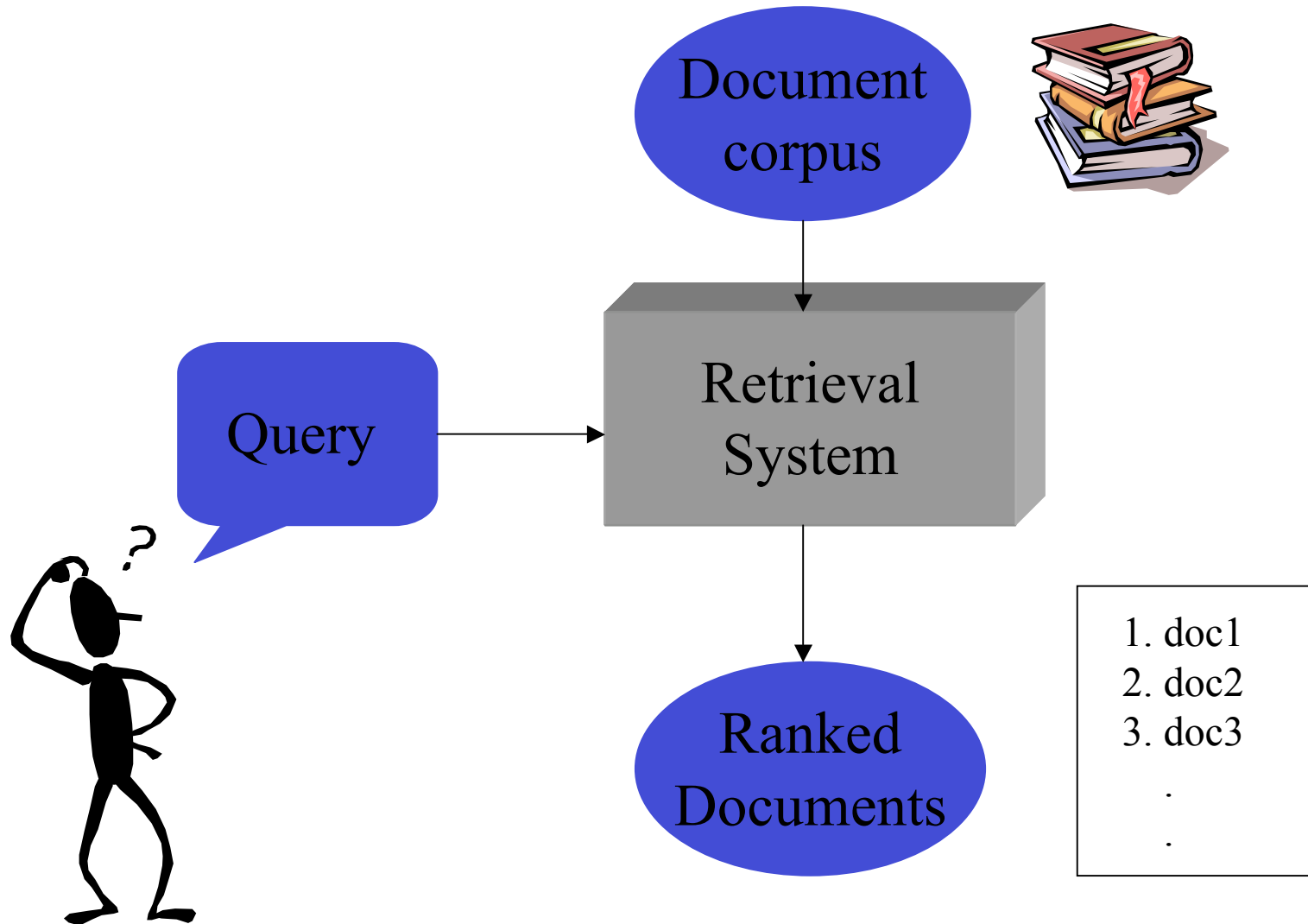
But I will talk about “Aggregation”

Mounia Lalmas  
mounia@acm.org

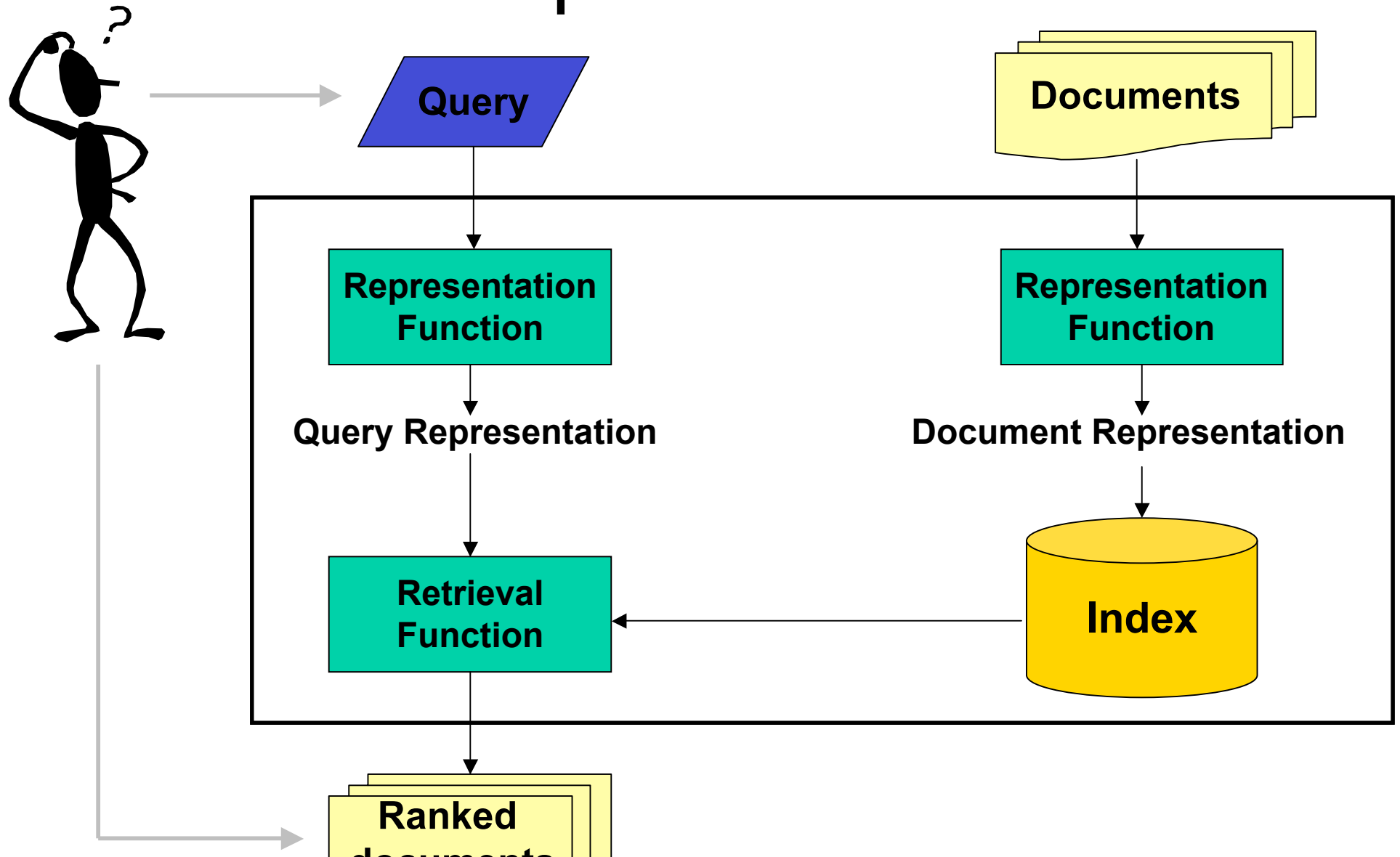
# Some sort of outline

- Document retrieval, information retrieval, web search
- Focused retrieval
- Aggregation
- Current work - Digest page

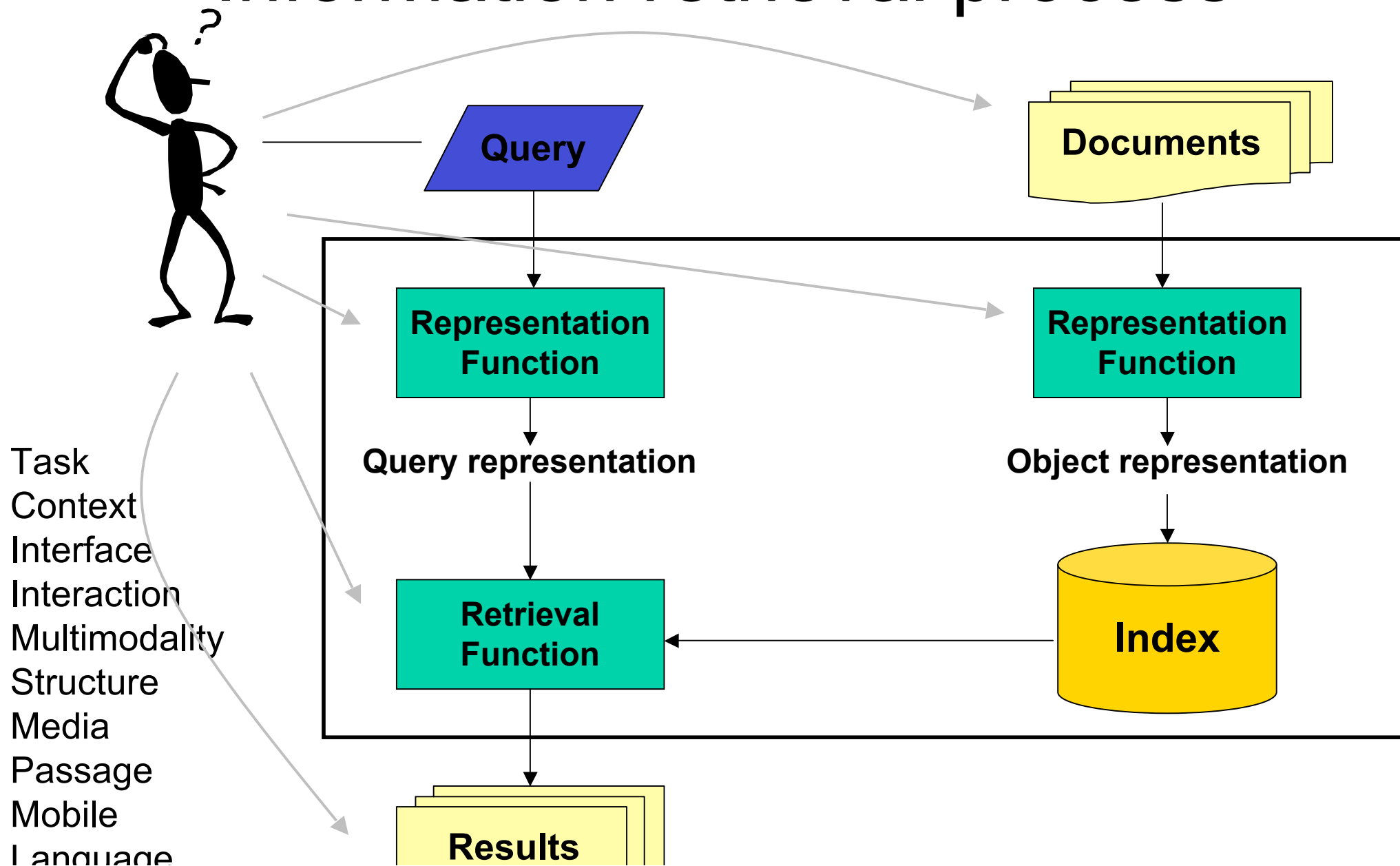
# Classical document retrieval



# Classical document retrieval process



# Information retrieval process



# Information retrieval and web search

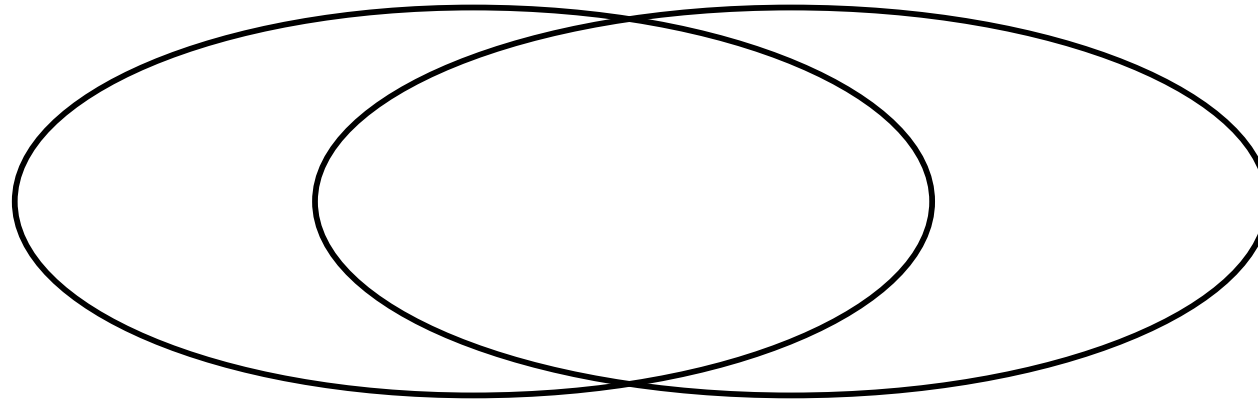
Google™

msn.™

YAHOO!®

web search

information retrieval



# Focused Retrieval - Question & Answering

## AnswerBus

When was Madonna born?

Ask

Type in your question in English, French, Spanish, German, Italian or Portuguese.

### Question:

When was Madonna born?

Possible answers: [XML](#) [TXT](#)

- [Madonna was born Madonna Louise Ciccone in Bay City, Michigan on August 16, 1958. Madonna was the third of eight children born to Madonna Fortin and Sylvio Ciccone.](#)
- [Madonna \(read more\) 5,958,624 plays scrobbled on Last.fm Madonna Ritchie \(born Madonna Louise Ciccone on August 16, 1958 in Bay City, Michigan, USA\) is an iconic pop singer, dancer, songwriter, producer, author and actress.](#)
- [Madonna Madonna was born Madonna Louise Veronica Ciccone in Bay City, Michigan, August 16, 1958, and grew up in a working-class household in Detroit.](#)
- [Madonna was born Madonna Louise Ciccone in Bay City, Michigan, in 1958.](#)
- [knowledgebroker -- Madonna was born as Madonna Louise Ciccone in Bay City, Michigan on August 16th, 1958.](#)
- [Madonna Louise Ciccone was born on August 16, 1958, in Michigan, to Sylvio \[a.k.a. Tony\] Ciccone and Madonna Fortin.](#)
- [Madonna, born Madonna Louise Veronica Ciccone in 1958, was the third of eight children.](#)
- [Madonna \(born Madonna Louise Ciccone in 1958\), an American pop icon, singer, and actress.](#)
- [Originally performed by Madonna \(born Madonna Louise Ciccone 8/16/58 in Bay City Michigan\)](#)
- [Madonna Wayne Gacy, aka Pogo \(born 1964\), keyboardist of rock band Marilyn Manson.](#)

# Focused Retrieval - XML Element Retrieval

dbdk\_training in Baseline System

 Search

query was: text classification naive bayes  
Results 1 - 10 of 100.  
Result pages: 1 2 3 4 5 6 7 8 9 10 next



## Search Result

**1: (0.247) Scalable Feature Mining for Sequential Data**

*Neal Lesh Mitsubishi Electric Research Lab Mohammed J. Zaki Rensselaer Polytechnic Institute Mitsunori Ogiwara University of Rochester*

**Result path:** [/article\[1\]/bdy\[4\]/sec\[5\]](#)

**2: (0.204) Probability and Agents**

*Marco G. Valtorta University of South Carolina mgv@cse.sc.edu Michael N. Huhns University of South Carolina huhns@sc.edu*

**Result path:** [/article\[1\]/bdy\[4\]/sec\[3\]](#)

**3: (0.176) Combining Image Compression and Classification Using Vector Quantization**

*Karen L. Oehler Member IEEE Robert M. Gray Fellow IEEE*

**Result path:** [/article\[1\]/bdy\[4\]/sec\[4\]/ss1\[2\]/ss2\[4\]](#)

**4: (0.175) Text-Learning and Related Intelligent Agents: A Survey**

*Dunja Mladenic J. Stefan Institute*

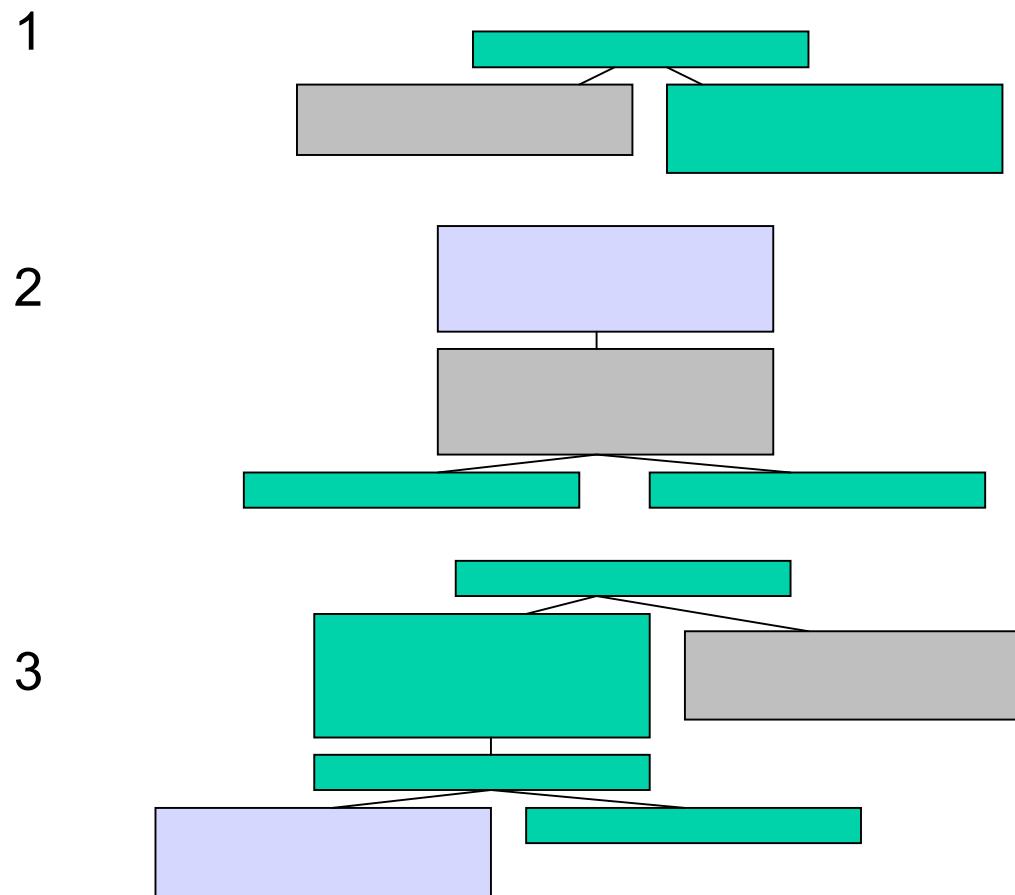
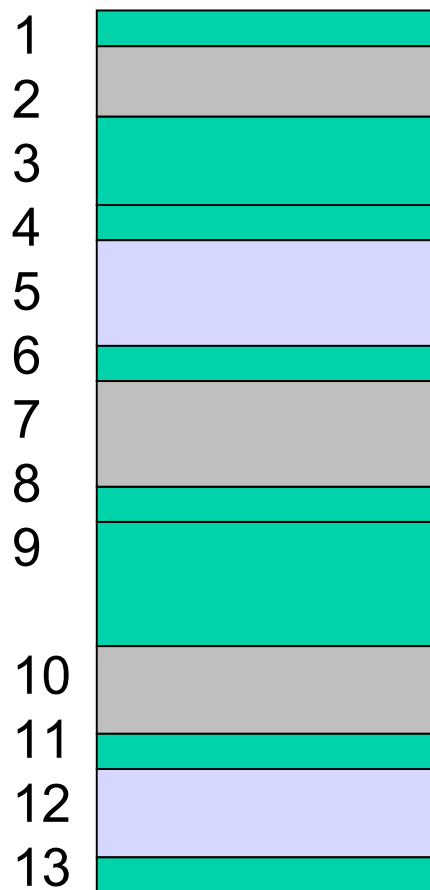
**Result path:** [/article\[1\]/bm\[5\]/app\[4\]/sec\[5\]](#)

**5: (0.175) Detecting Faces in Images: A Survey**

*Ming-Hsuan Yang Member IEEE David J. Kriegman Senior Member IEEE Narendra Ahuja Fellow IEEE*

**Result path:** [/article\[1\]/bdy\[4\]/sec\[2\]/ss1\[9\]/ss2\[10\]](#)

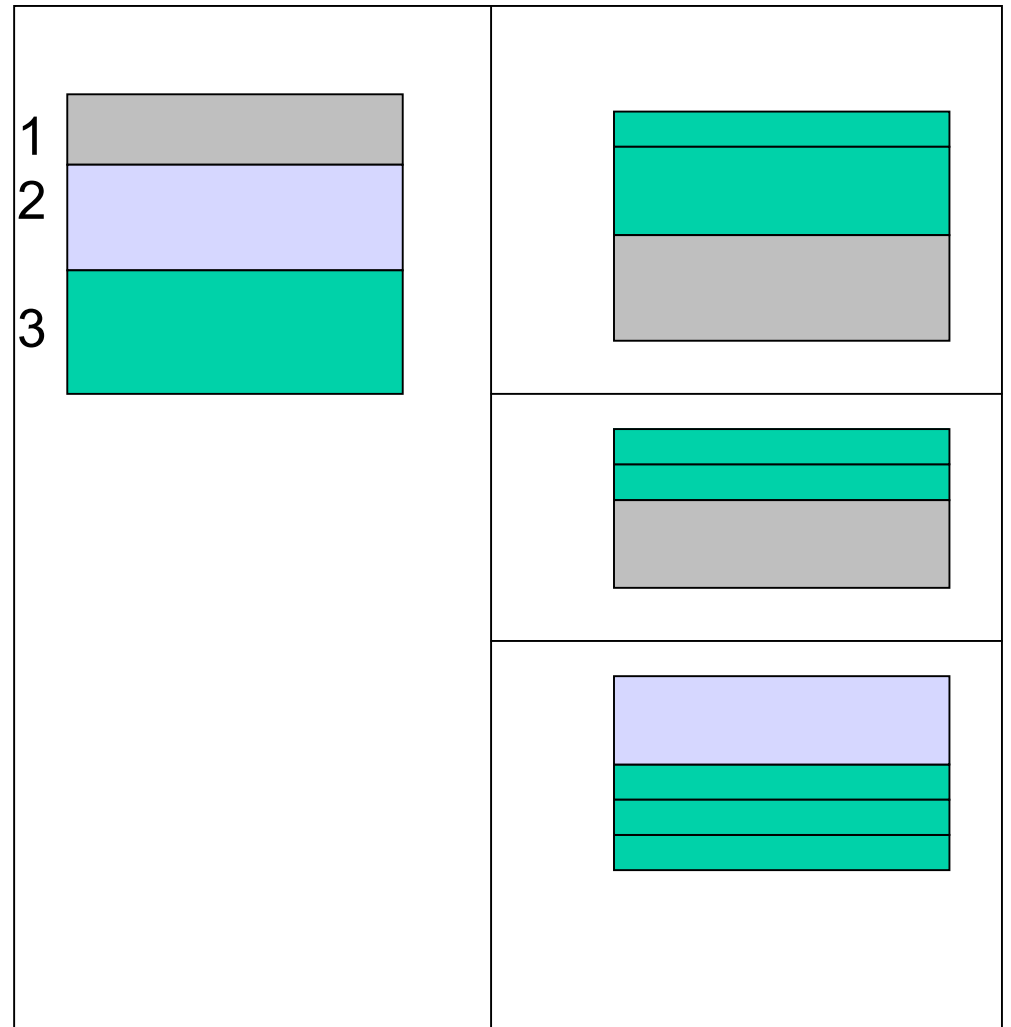
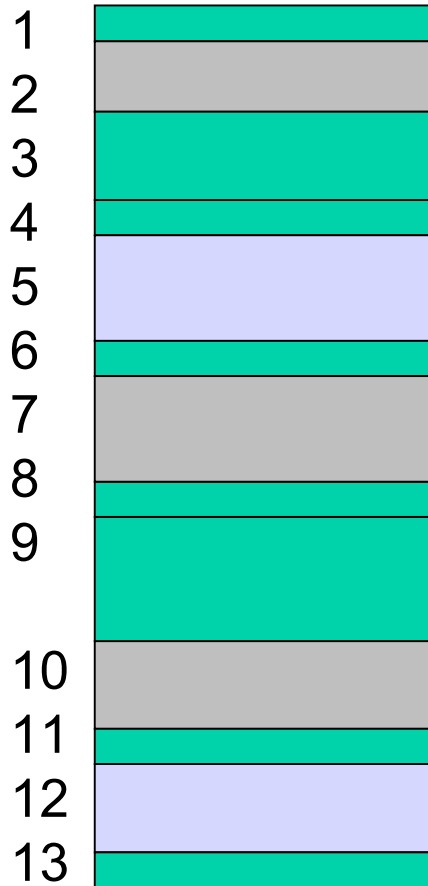
# Aggregated (virtual) documents



# Aggregated (virtual) documents

- Web search
  - Clustering (Clusty)
  - Summarisation (WebInEssence)
  
- News domains
  - Topic detection & tracking (TREC track)

# Aggregated views



# Aggregated views

alpha<sup>™</sup>  
beta

Hiya Guest | Sign in

Customise my search

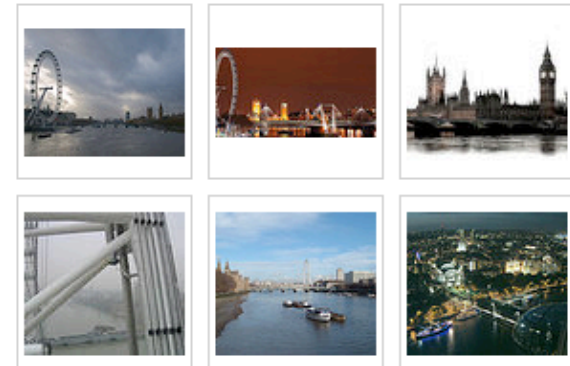
Also try: hellgate london, london time, london hotels, lauren london, london eye more...

Web Search by Yahoo!7 Search

782,000,000 results

1. [Visit London - London's official website - London Tourist ...](#)  
Visit **London** with ease using **London** Tourist, Entertainment and Travel Guides. ... those tiaras and feather boas – the world's campest musical is back in **London!** ...  
<http://www.visitlondon.com/>
2. [LondonTown.com | The Number One Internet Site for London England ...](#)  
Sightseeing, attraction and other tourist information. Theater tickets, **London** maps and more. ... Christmas Lights in **London** 2007 ...  
<http://www.londontown.com/>
3. [London Entertainment Guide from The Evening Standard | This is London](#)  
... tickets booking for **London** covering Film, Restaurants, ... **London** 'safest place' ... This is **London** Magazine Ltd. Daily Mail Mail on Sunday Travel Mail ...  
<http://www.thisislondon.co.uk/>
4. [London - Wikipedia, the free encyclopedia](#)  
An important settlement for two millennia, **London** is one of the world's leading ... **London** is a major tourist destination with four world heritage sites, several ...  
<http://en.wikipedia.org/wiki/London>
5. [Mayor of London, the London Assembly and the Greater London Authority](#)  
The official website for the Mayor of **London**, the **London** Assembly and the Greater **London** Authority ... **London** Assembly members, Richard Barnes and Tony Arbour, ...  
<http://www.london.gov.uk/>
6. [Time Out London - Discover Things To Do in London Including ...](#)  
Highlights this month's events in **London** and provides details of accommodation, ... **London** Jazz Festival ... **London** ice-skating rinks. Start practising your ...  
<http://www.timeout.com/london/>
7. [BBC - London - London Homepage](#)




Photos by Flickr



« More results

« Prev | Next »

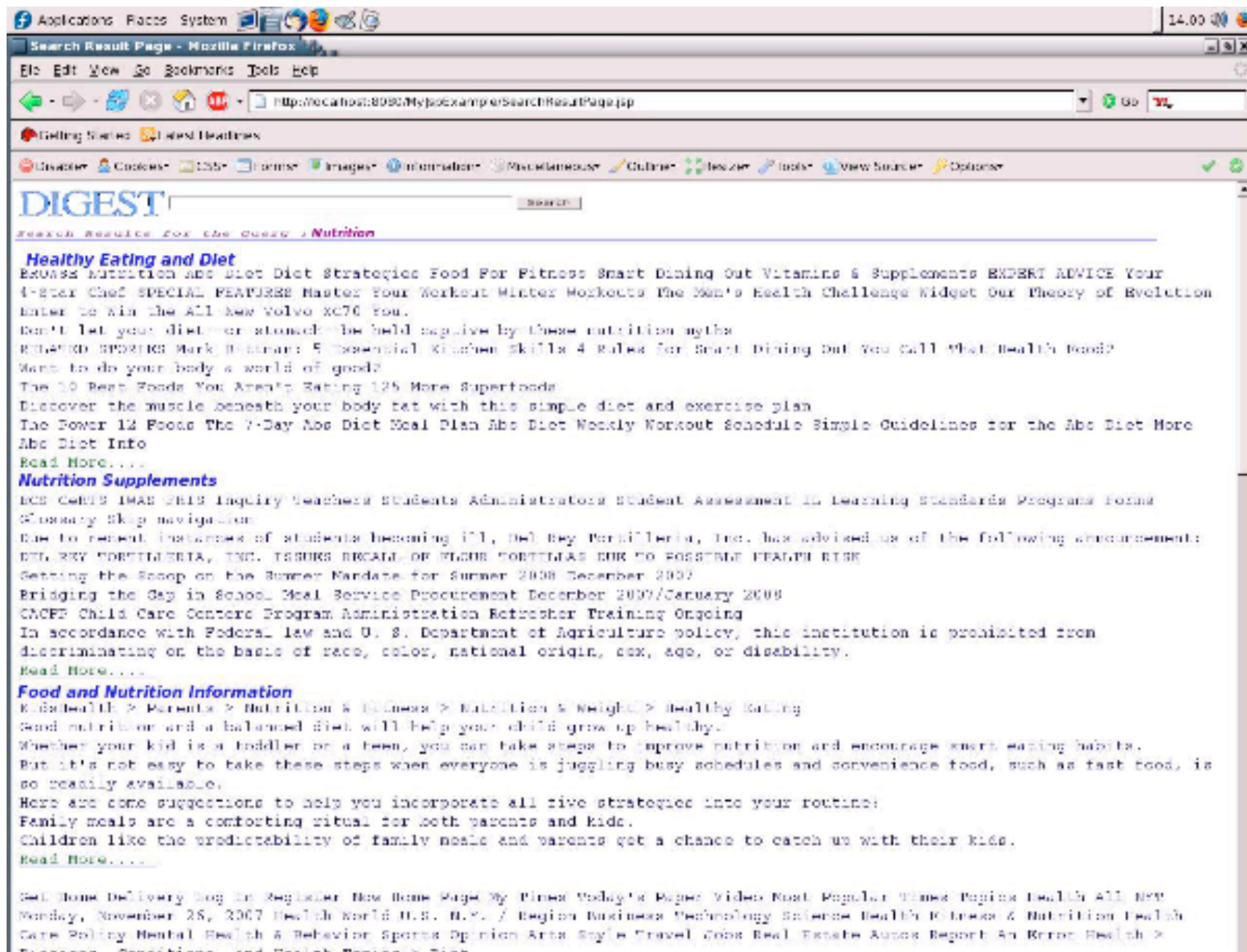
Yahoo!7 Answers

1.  [London!!!!!!!!!!?](#)  
...northwest of England and i love going down to London a few times a yr, & each time i go...  
<http://au.answers.yahoo.com/question/?qid=2007C>
2.  [London?????????](#)  
I am going London tomorrow and im a kid what places are main...  
<http://au.answers.yahoo.com/question/?c>
3.  [london....?](#)  
...of beautiful people down the place i live in london getting on and off the bus...do any...  
<http://au.answers.yahoo.com/question/?qid=20061>

# Current work

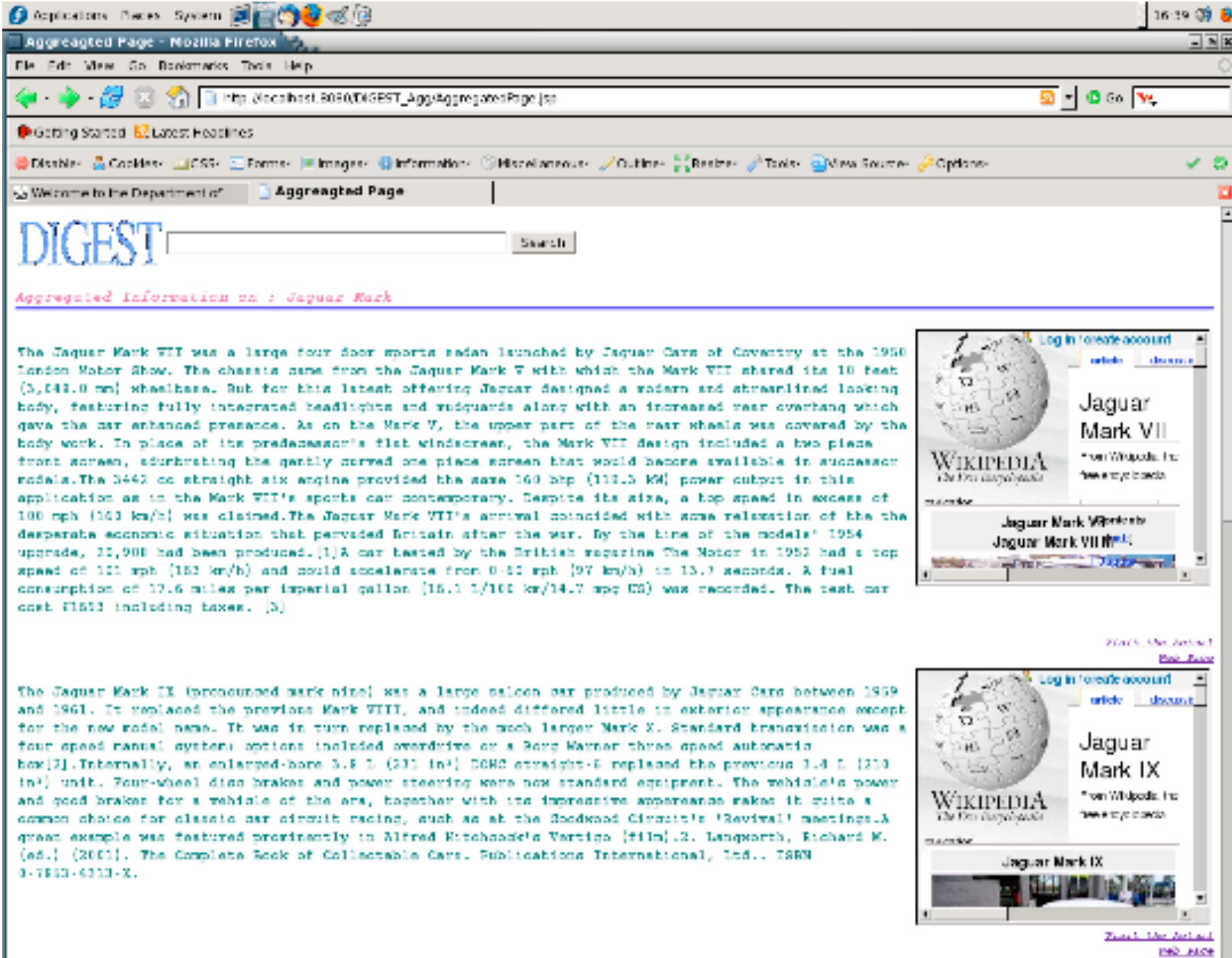
- Digest page
  - Yahoo! Wikipedia
  - Clustering (Carrot<sup>2</sup>)
  - Multi-document summarization
- Ranking of digest pages
- Aggregated digest page
  - Query context

# Current work - ranked list of digest pages



The screenshot shows a Mozilla Firefox browser window with the address bar displaying "http://localhost:8080/MyJspExample/SearchResultPage.jsp". The page title is "Search Result Page - Mozilla Firefox". The browser's toolbar includes buttons for Back, Forward, Home, Stop, and Reload, along with a search bar and a "Go" button. Below the browser window, the page content is displayed. The main heading is "DIGEST" in large blue letters. Below it, there is a search bar and a link to "Search Results for the query: Nutrition". The page is divided into several sections, each with a blue heading and a list of links. The first section is "Healthy Eating and Diet", which includes links like "EXCESS Nutrition and Diet Diet Strategies Food For Fitness Smart Dining Out Vitamins & Supplements EXPERT ADVICE Your 4-Star Chef SPECIAL FEATURES Master Your Workout Winter Workouts The Men's Health Challenge Kidset Our Theory of Evolution Enter to Win the All-New Volvo XC70 You." and "Don't let your diet or stomach be held hostage by these nutrition myths". The second section is "Nutrition Supplements", which includes links like "ECS CEHS IMAS PBS Inquiry Teachers Students Administrators Student Assessment IS Learning Standards Programs Form" and "Due to recent instances of students becoming ill, DEL Rey Tortilleria, Inc. has advised us of the following announcement: DEL REY TORTILLERIA, INC. ISSUES RECALL OF FLOUR TORTILLAS DUE TO POSSIBLE HEALTH RISK". The third section is "Food and Nutrition Information", which includes links like "KidsHealth > Parents > Nutrition & Fitness > Nutrition & Weight > Healthy Eating" and "Good nutrition and a balanced diet will help your child grow up healthy."

# Current work - presenting a digest page



The screenshot shows a Mozilla Firefox browser window displaying a 'DIGEST' page. The browser's address bar shows the URL 'http://localhost:8080/DIGEST\_AggrAggregatedPage.jsp'. The page features a search bar with the word 'DIGEST' in large blue letters and a 'Search' button. Below the search bar, there is a section titled 'Aggregated Information on : Jaguar Mark'. The main content area contains two paragraphs of text describing the Jaguar Mark VII and Jaguar Mark IX. To the right of the text, there are two preview cards for Wikipedia articles: 'Jaguar Mark VII' and 'Jaguar Mark IX'. Each card includes a globe icon, the Wikipedia logo, and a 'Log in / create account' link. The browser's status bar at the bottom shows the page title 'Aggregated Page' and the URL.

Applications | Files | System | 16:39

Aggregated Page - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost:8080/DIGEST\_AggrAggregatedPage.jsp

Getting Started | Latest Headlines

Disable Cookies CSS Forms Images Information Miscellaneous Outline Reader Tools View Source Options

Welcome to the Department of | Aggregated Page

# DIGEST

Aggregated Information on : Jaguar Mark

The Jaguar Mark VII was a large four door sports sedan launched by Jaguar Cars of Coventry at the 1950 London Motor Show. The chassis came from the Jaguar Mark V with which the Mark VII shared its 10 feet (3,048.0 mm) wheelbase. But for this latest offering Jaguar designed a modern and streamlined looking body, featuring fully integrated headlights and mudguards along with an increased rear overhang which gave the car enhanced presence. As on the Mark V, the upper part of the rear wheels was covered by the body work. In place of its predecessor's flat windscreen, the Mark VII design included a two piece front screen, adorning the gently curved one piece screen that would become available in successor models. The 3442 cc straight six engine provided the same 160 bhp (119.3 kW) power output in this application as in the Mark VII's sports car contemporary. Despite its size, a top speed in excess of 100 mph (163 km/h) was claimed. The Jaguar Mark VII's arrival coincided with some relaxation of the the desperate economic situation that pervaded Britain after the war. By the time of the models' 1954 upgrade, 11,908 had been produced. [1] A car tested by the British magazine The Motor in 1953 had a top speed of 121 mph (195 km/h) and could accelerate from 0-60 mph (97 km/h) in 13.7 seconds. A fuel consumption of 17.6 miles per imperial gallon (16.1 l/100 km/14.7 mpg US) was recorded. The test car cost £1553 including taxes. [3]

The Jaguar Mark IX (pronounced mark nine) was a large saloon car produced by Jaguar Cars between 1959 and 1961. It replaced the previous Mark VIII, and indeed differed little in exterior appearance except for the new model name. It was in turn replaced by the much larger Mark X. Standard transmission was a four speed manual; system options included overdrive or a Borg Warner three speed automatic box [2]. Internally, an enlarged-bore 3.4 L (211 in³) DOHC straight-6 replaced the previous 3.4 L (210 in³) unit. Four-wheel disc brakes and power steering were now standard equipment. The vehicle's power and good brakes for a vehicle of the era, together with its impressive appearance makes it quite a common choice for classic car circuit racing, such as at the Goodwood Circuit's 'Revival' meetings. A green example was featured prominently in Alfred Hitchcock's Vertigo [film]. 2. Langworth, Richard M. (ed.) (2002). The Complete Book of Collectable Cars. Publications International, Inc.. ISBN 0-7853-6333-X.

Log in / create account  
article discuss

## Jaguar Mark VII

From Wikipedia, the free encyclopedia

Jaguar Mark VII  
Jaguar Mark VII (left)

Log in / create account  
article discuss

## Jaguar Mark IX

From Wikipedia, the free encyclopedia

Jaguar Mark IX

Log in / create account  
article discuss

# Current work - context of query

The screenshot shows a web browser window with the address bar containing 'http://www.google.com/search?q=land+rover&btnG=Google+Search'. The page displays search results for 'Land Rover'. The main content area features a snippet from Wikipedia titled 'Aggregated Information from Yahoo! Wikipedia on : Land Rover'. The snippet text reads: 'land rover is an all-terrain vehicle and multi purpose vehicle (mpv) manufacturer, based in oxley, england, uk with two uk production plants at malvern, near wolverhampton and solihull, west midlands and solihull, westmidlands. Originally the term land rover referred to one specific vehicle, a pioneering civilian all-terrain utility vehicle launched on April 29, 1948, at the Amsterdam Motor Show, but was later used as a brand for several distinct models, all four wheel drive. Starting out as part of The Rover Car Company or Rover, Land Rovers were designed and manufactured as a range of four wheel drive vehicles under a succession of owners, including British Leyland, British Aerospace and BMW. In 2001, the marque was sold by BMW to the Ford Motor Company, becoming part of their Premier Automotive Group. Media reports stated that Jaguar and Land Rover became owned by the Indian company Tata Motors on March 21, 2008, (It is understood that Ford Motor Company Ltd. will not retain any shareholding in either the Jaguar or Land-Rover companies, unlike Aron Martin whom on its sale a small shareholding was retained; the total sum to be paid in cash by Tata Motors is approximately us \$ 2.5 billion, Ford will then contribute up to us \$10 million to the Jaguar Land Rover portfolio).

 To the right of the text is a small image of a globe with the text 'Land Rover' and 'From Wikipedia, the free encyclopedia'. Below the text is a small image of a Land Rover vehicle. At the bottom of the main content area is a row of small thumbnail images. On the right side of the browser window, there are two columns of search results. The top column is titled 'More results from Yahoo! on :Land Rover' and contains several links, including 'Land Rover Dealer UK - Parts and more Land Rover Range Rover Dealer in Southdownshire - also used cars and vans in Southdownshire ...', '2007 Land Rover Range Rover Dealer, 2008 Land Rover Range Rover Dealer, 2007 Land Rover Range Rover Sport Dealer, 2008 Land Rover Range Rover Sport Dealer ...', 'Land sells Land Rover, Jaguar', and 'BBC Northwest News Talk Radio ... Land sells Land Rover, Jaguar. Last Updated 3/28/2008 14:55:00 AM ... "Jaguar and Land Rover are terrific brands," said Alan ...'. The bottom column is titled 'More from Yahoo! on :Land Rover' and contains links such as 'Rover Land Rover Discovery - facelift', 'Rover Land Rover Discovery - facelift. Arrive in Rover Land Rover Discovery - facelift. Arrive in Rover Land Rover Discovery. Le principali modifiche antiche sono concentrate sui fessoni parziali anteriori e posteriori radioganti e in tutta carrozzeria.', and 'Land Rover Freelander 2 voted Best Small SUV 2004'.

# Ask X - aggregated search

The screenshot shows a Mozilla Firefox browser window displaying the Ask.com search results for the query "jaguar". The browser's address bar shows the URL: <http://uk.ask.com/web?q=jaguar&search&dm=all&qspr=05a-11261-din&chid=...>

The search results are organized into several sections:

- Search Bar:** The search term "jaguar" is entered in the search box. Below it, there are options for "The Web" and "UK Only", and a "Advanced" link.
- Narrow Your Search:** A list of related terms including "Jaguar Cars", "Jaguar Animal", "Jaguar Owners Club", "Jaguar Science", "Jaguar Gifts", "Jaguar UK", "Jaguar Parts", "Jaguar X Type", "Jaguar XJ600", and "Jaguar History".
- Expand Your Search:** A list of car brands including "Audi", "BMW", "Aston Martin", "Mercedes", "Ferrari", "Lamborghini", and "Audi Trade".
- Main Results:**
  - Jaguar (Animal):** A result for the big cat, including a small image and text describing it as the biggest and most powerful North American cat. It lists its taxonomic classification: Kingdom: Animalia, Phylum: Chordata, Class: Mammalia, Order: Carnivora, Family: Felidae, Genus: Panthera, Species: Panthera onca. It also includes a "Key Facts" link.
  - Jaguar - Official Site:** A sponsored result with the text "Discover the luxury experience that is Jaguar." and the URL [www.jaguar.co.uk](http://www.jaguar.co.uk).
  - Jaguar Cars:** A result with the text "Find your Jaguar Online today. View Multiple Photos & Exclusive Details" and the URL [www.lookerjaguar.co.uk](http://www.lookerjaguar.co.uk).
  - Jaguars available to buy:** A result with the text "Yorkshire dealership. Large stock online. Free national delivery." and the URL [www.spryward-jaguar.co.uk](http://www.spryward-jaguar.co.uk).
  - New Jaguar Range:** A result with the text "Charles Hurst Jaguar Range Online Exclusive Offers - Enquire Now!" and the URL [www.charleshurstjaguar.co.uk](http://www.charleshurstjaguar.co.uk).
  - Jaguar:** A result with the text "Buy your New Jaguar at a discount Great Choice, Super Savings" and the URL [www.laplayhercars.co.uk/jaguar](http://www.laplayhercars.co.uk/jaguar).
  - Jaguar US - Home:** A result with the text "Jaguar USA official website..." and the URL [www.jaguarusa.com/](http://www.jaguarusa.com/).
- News:** A section titled "News" showing "Top news for Views" with links to "Autism cases" (24 hours ago) and "Drop in UK domestic M&A activity" (24 hours ago).
- Video:** A section titled "Video" showing "Top Marquee: Jaguar Part 2" (04:28 - Discovery) and "Top Marquee: Jaguar Part 2" (00:28 - Discovery).
- Feeds:** A section titled "Feeds" showing links for "Jaguar", "Jaguar Marketing System", and "Optical Jaguars Retro Smur".

# Aggregated “answers”

## Why?

- reduce information overload (is there one?)
- increase result space (is this needed?)
  
- provide/capture context/focus (country, language, device, literacy, ...)
- presentation and interface (expectation for Google type interface very high! Is it the case?)
- cross/multi-media content (refer to StoryBank)
- cross/multi-lingual content (is there a need?)

*Is the above too complicated? Not needed? Not feasible? In 10 years?*