

# Internet Search Techniques

Finding What You Want on the Web Easily,  
Quickly, and (sort of) Effortlessly...

Presented by: Sharon Coward



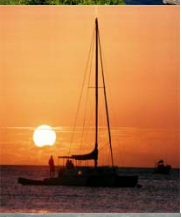


# Overview

## Objectives:

- Understand the Web as a repository of information.
- Explore different search tools.
- Learn to use the tools appropriately
- Evaluate search results.

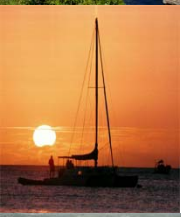




# Internet Search Technologies

- Internet = network of computers
- The Web = one of the services available via the Internet; interconnected documents & other resources, linked by hyperlinks and URLs

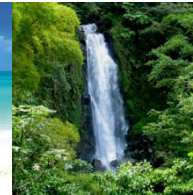
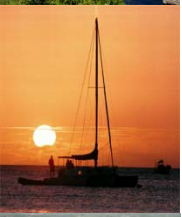




# The Web – how big is it?

- Google: 5 million terabytes = 5 trillion megabytes of data.
- Google indexes only 200 terabytes i.e. 0.004%
- 2005 – 11.5 billion pages indexable web; doubles in size every 5 yrs.





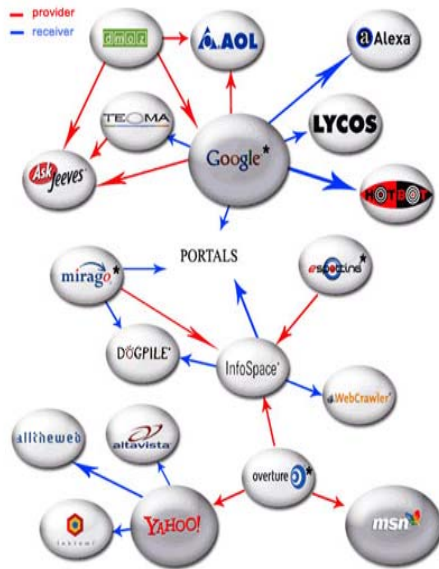
# Search Tools

- Search Engines
- Meta-search Engines
- Information Gateways
- Invisible/Deep Web





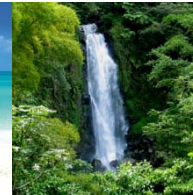
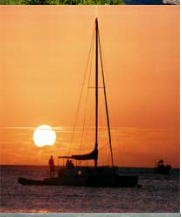
# What is a “Search Engine”?



- 1. A (computer) program that searches documents for specified keywords and returns a list of the documents where the *keywords* were found.
- Often used to specifically describe systems like Google and Bing that enable users to search for documents on the World Wide Web

<http://www.webopedia.com>





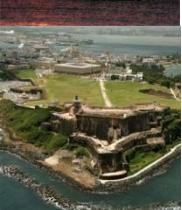
# Search Engines

- Number of pages searched can vary
- Good results depend on using proper search syntax not just the scope of the engine's coverage
- **Good For:**

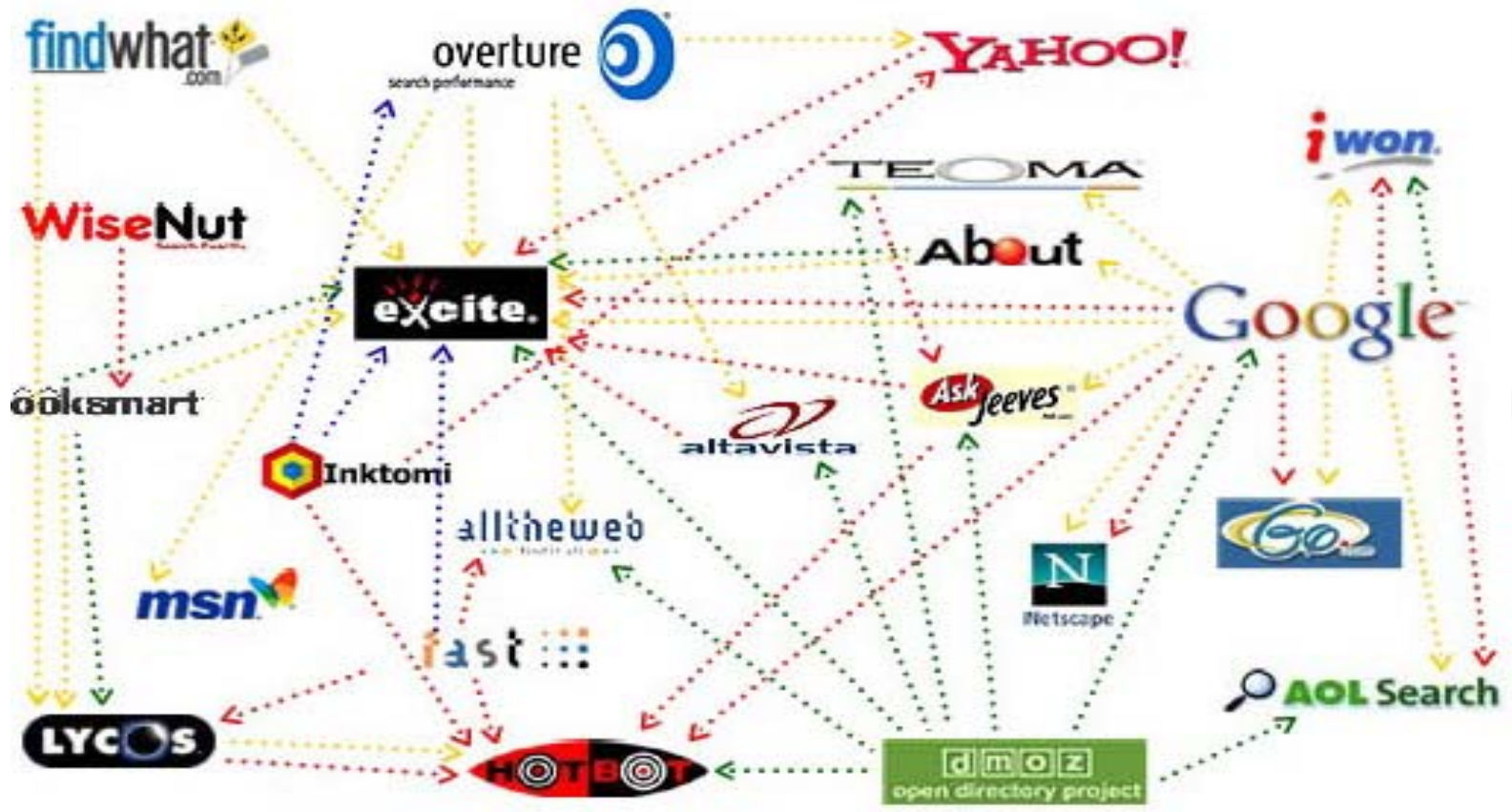
well defined topics to search; looking for specific sites; want a large number of websites returned for topic; retrieve particular types of documents, eg. Pdf

**Not Good For:** Browsing through a subject area.































# Search Engines



# Search Engine Relationship

Search Engine	Web Search Results	Sponsored Listings	Directory Listings
			N/A
			N/A
			N/A
			N/A
			N/A
		N/A	
			N/A
			
			N/A
			N/A
			N/A
			N/A
			





# Meta Search Engines

- Skim-search several search engines at once
- Usually reach about 10% of results of each engine they visit
- Cannot perform advanced-style searches which use engine-specific syntax
- **Good For:**  
quick search engine results overview, doing simple searches with 1 or 2 keywords; want a small # relevant results; problems finding what you want; convenient to search different content sources from one page
- **Not Good For:** comprehensive results from a complex search



# Meta-Search Engines

**dogpile**

Web | Images | Video | News | Yellow Pages | White Pages

Go Fetch!

Google Yahoo! bing Ask Advanced Search Preferences

**Favorite Fetches**

- [Nobel Peace Prize](#)
- [Christmas Play Lists](#)
- [911 Elephant](#)
- [Holiday Gift Guides](#)
- [Woods Mistress List](#)
- [Cheap Airline Tickets](#)

**Quick Starts**

- [Add Dogpile To Your Browser](#)
- [Download Toolbar](#)
- [Add Dogpile To Your Site](#)
- [SearchSpy](#)
- [Bookmark and Share](#)
- [Horoscope](#)

**metacrawler**

SEARCH THE SEARCH ENGINES!

Including Google, Yahoo!, Bing and Ask. [Learn More](#)

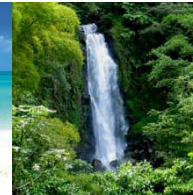
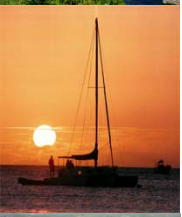
Web | [Images](#) | [Video](#) | [News](#) | [Yellow Pages](#) | [White Pages](#)

SEARCH

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

Dogpile - [www.dogpile.com](http://www.dogpile.com); Metacrawler - [www.metacrawler.com](http://www.metacrawler.com)





# Meta Search Engines

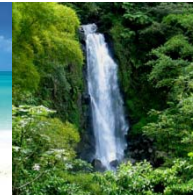
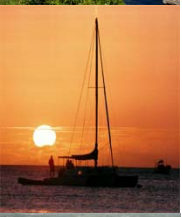


Results: 100 ▾      SortBy: Relev ▾      Summaries:      

- [LookAhead™ ...](#)  
Dynamic Search Tools
- [RSS Feed Search ...](#)  
Dynamically by Title
- [WikiWax ...](#)  
Wikipedia Search
- [Nextaris ...](#)  
Your Web Dashboard
- [News Accumulator ...](#)  
50,000 News Topics
- [ShopEasier ...](#)  
Shopping 'discovery'

SurfWax – [www.surfwax.com](http://www.surfwax.com)





# Meta Search Engines

copernic

Web Images Audio Video News

- Copernic - <http://find.copernic.com>





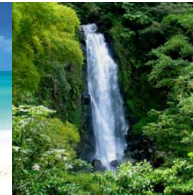
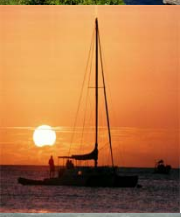
# Information Gateways

- Subject directories, virtual libraries
- Compiled by people, not robots
- Subject categories
- More focus on sifting for relevance and quality

Good For: you have a clear topic but not unique keywords; browse for ideas

Not Good For: Quickly finding information from widely varying themes





# Information Gateways

## YAHOO! SEARCH

[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Directory](#) | [more »](#)

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

[\[ Show Editorial Features \]](#)

[Browse Yahoo!'s categorized guide to the Web.](#)



[Arts & Humanities](#)

[Business & Economy](#)

[Computers & Internet](#)

[Education](#)

[Entertainment](#)

[Government](#)

[Health](#)

[News & Media](#)

[Recreation & Sports](#)

[Reference](#)

[Regional](#)

[Science](#)

[Social Science](#)

[Society & Culture](#)

Other:

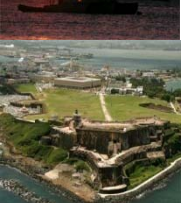
[New Additions](#)

[RSS Feeds](#)

Copyright © 2009 Yahoo! All rights reserved. [Privacy](#) / [Legal](#) - [Submit Your Site](#)

# http://dir.yahoo.com/





# Information Gateways



[Web](#) [Images](#) [Groups](#) [News](#) [Shopping](#) [Maps](#) [Scholar](#) [more »](#)

Search Directory

[Preferences](#)  
[Directory Help](#)

The web organized by topic into categories.

## Arts

[Movies](#), [Music](#), [Television](#), ...

## Business

[Industrial Goods and Services](#), [Finance](#), [Jobs](#), ...

## Computers

[Hardware](#), [Internet](#), [Software](#), ...

## Games

[Board](#), [Roleplaying](#), [Video](#), ...

## Health

[Alternative](#), [Fitness](#), [Medicine](#), ...

## World

[Deutsch](#), [Español](#), [Français](#), [Italiano](#), [Japanese](#), [Korean](#), [Nederlands](#), [Polski](#), [Svenska](#), ...

## Home

[Consumers](#), [Homeowners](#), [Family](#), ...

## News

[Media](#), [Newspapers](#), [Current Events](#), ...

## Recreation

[Food](#), [Outdoors](#), [Travel](#), ...

## Reference

[Education](#), [Libraries](#), [Maps](#), ...

## Regional

[Asia](#), [Europe](#), [North America](#), ...

## Science

[Biology](#), [Psychology](#), [Physics](#), ...

## Shopping

[Autos](#), [Clothing](#), [Gifts](#), ...

## Society

[Issues](#), [People](#), [Religion](#), ...

## Sports

[Basketball](#), [Football](#), [Soccer](#), ...

Google Directory - [www.google.com/dirhp](http://www.google.com/dirhp)





# Information Gateways

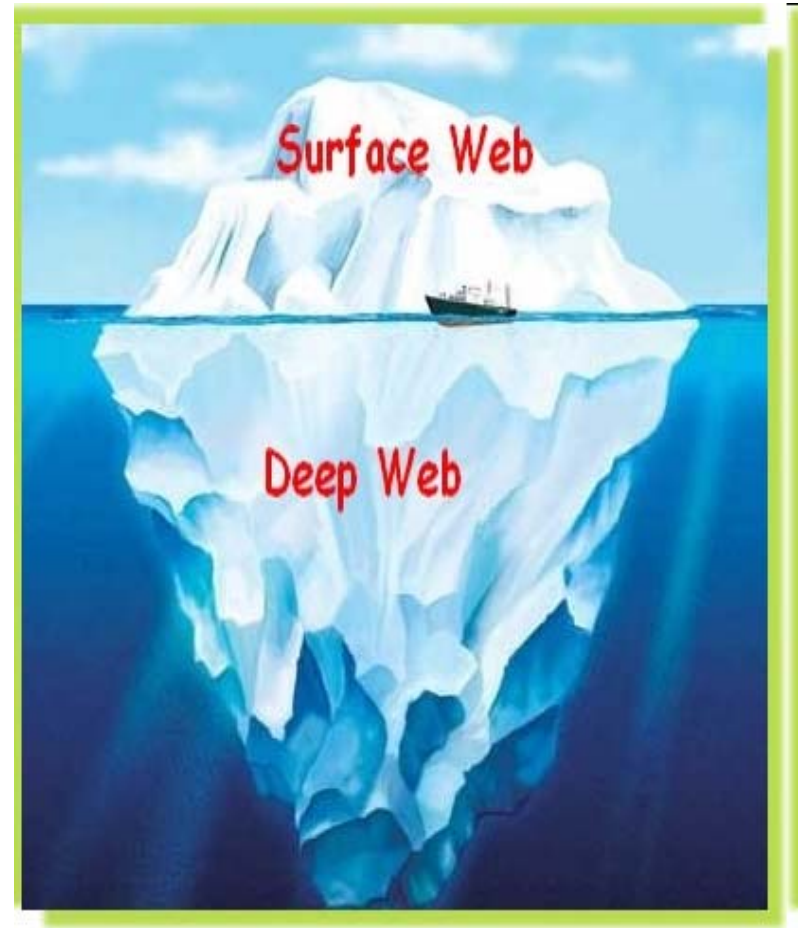
- Questia.com - <http://www.questia.com/Index.jsp> (full text online library >70,000 books)
- ELDIS: the Gateway to Development Information  
<http://www.eldis.org/> (4000 sites)
- Open Learn  
<http://openlearn.open.ac.uk/> (open university course materials)
- Open Directory Project  
<http://www.dmoz.org> (largest human edited directory)





# Invisible/Deep Web

- 91,000 terabytes vs 167 terabytes in surface Web
- Search engines can't access content – databases, non-text files; password protected areas; dynamic content





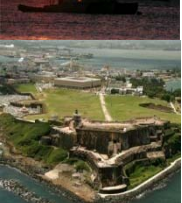
# Invisible/Deep Web

- Dynamic content: - returned in response to a submitted query or accessed only through a form,
- Unlinked content: pages which are not linked to by other pages
- Private Web: sites that require registration and login (password-protected resources).
- Searchable - Entry pages can be found using other search tools; include term 'database' in search

Good For: Gathering specific kinds of data

Not Good For: Browsing through a subject area





# Invisible/Deep Web

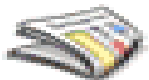
[Web](#) [Images](#) [Maps](#) [News](#) [Video](#) [Gmail](#) [more](#) ▼

# Google™

"global warming"

[Web](#) [News](#) [Video](#) [Books](#) [Images](#)

## News results for "global warming"



[Global-Warming Skeptics Raise A Storm](#)  
 By Nikola Krastev NEW YORK -- These ...  
 warming. Two years ago, an international ...  
[RadioFreeEurope/RadioLiberty](#) - 319 related ...  
[Prince highlights global warming](#) - The ...  
[Cold reality of global warming efforts](#) -

## [Global warming](#) - Wikipedia, the free encyclopedia

Global warming is the increase in the average temperature of the Earth's atmosphere and the oceans since the mid-twentieth century and its projected ...  
[en.wikipedia.org/wiki/Global\\_warming](http://en.wikipedia.org/wiki/Global_warming) - 418k - [Cached](#) - [Similar pages](#)

- [Shopping](#)
- [Groups](#)
- [Books](#)
- [Scholar](#)
- [Finance](#)
- [Blogs](#)
- [YouTube](#)
- [Calendar](#)
- [Photos](#)
- [Documents](#)
- [Reader](#)
- [Sites](#)
- [even more](#) »

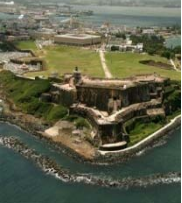
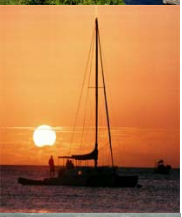




# Deep Web Expanding

- Blog postings
- Comments
- Discussions and other communicative activities on social networking sites
- Bookmarks and citations stored on social bookmarking sites





# Invisible/Deep Web

- [www.science.gov](http://www.science.gov) - over 40 databases, 1,950 selected websites, 200 million pages
- <http://www.deepwebtech.com/>
- <http://infomine.ucr.edu/>
- <http://www.completeplanet.com/> - 70,000+ databases
- <http://www.stumbleupon.com/> - compiled by humans since 2002

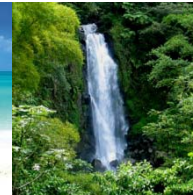




# Search Strategies

1. Identify your concepts
2. Make a list of search terms/ keywords for each concept e.g.
  - global warming/ greenhouse effect
  - greenhouse gases/ climate change
3. Specify the logical relationships among your search terms
4. Be specific – golden retriever vs dog





# Search Strategies

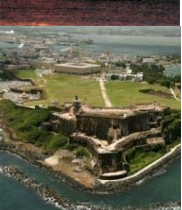
Google™

Google Search

I'm Feeling Lucky

[Advanced Search](#)  
[Language Tools](#)





# Advanced Search



Advanced Search

[Advanced Search Tip](#)

rain snow

Find web pages that have...

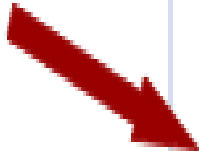
all these words:

this exact wording or phrase:

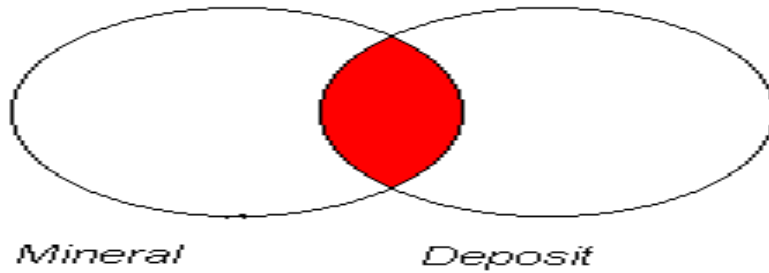
one or more of these words:

OR

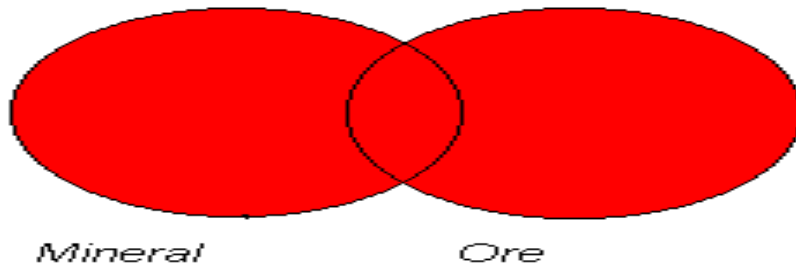
OR



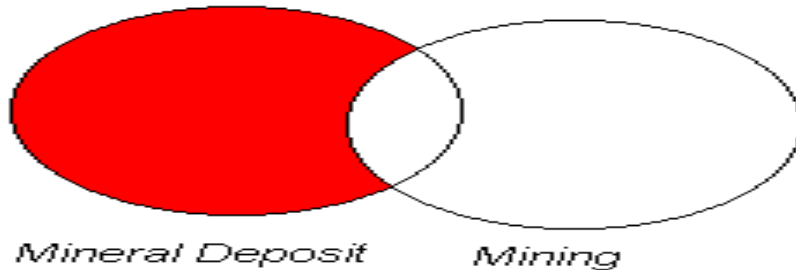
# Boolean Operators



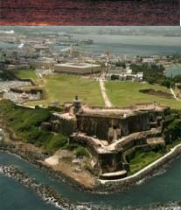
**AND** Articles will contain mineral and deposit



**OR** Articles will contain either mineral or ore



**NOT** Articles will contain mineral deposit but not mining





# Advanced Techniques

- Use quotation marks “...” to specify exact phrases:  
“internet marketing”
- Use the plus (+) and minus (-) sign to include and exclude words:  
“internet marketing +facebook” vs “internet marketing –facebook”





# Advanced Search

- Wildcard (\*). Searching for *'looking for \**' will return results that have the words **'looking for dogs'**, 'looking for cats' etc.
- Stop words ignored -“a, an,the, and”. Use a + or – symbol to force the engine to include those words in your search.



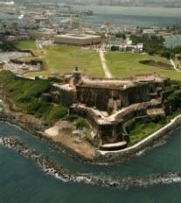
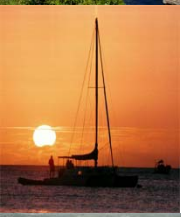


# Advanced Search

- ~ **Similar Words**

Search for similar words, or synonyms. Searching *search ~tips* will return results with 'help', 'guide', 'tutorial' etc.

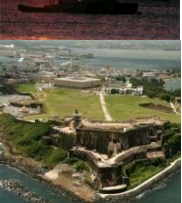




# Advanced Search

- Web page title  
*allintitle:jamaica*
- Website or domain  
*site:barbados.org "beach hotel"*





# Advanced Search

(e.g. youtube.com, .edu)

## [\[-\] Date, usage rights, numeric range, and more](#)

Date: (how recent the page is)

anytime

[Usage rights:](#)

not filtered by license

Where your keywords show up:

anywhere in the page

Region:

any region

Numeric range:

(e.g. \$1500..\$3000)

[SafeSearch:](#)

Off  On





# Advanced Search

- File type

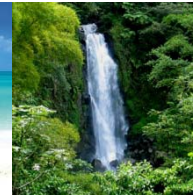
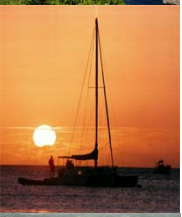
filetype:ppt site:edu "global warming"

- Definitions

define:pixel

define:"due diligence"





# Advanced Search

- **Truncation:** Searches on the root of the word adding different word endings or plurals.

*Educat\** searches *educator, education, educational, educated.*

*Colo\*r* would find documents that contain color and colour.





# Examine Results

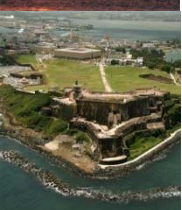
**Authority:** Who owns the site? Credentials?

.gov, .edu, .mil, .org usually more reliable than  
.com

**Currency:** How up-to-date is the information?

Check to see if you can tell when the last time  
the site was updated.





# Links

- List of search engines: <http://thesearchenginelist.com>
- Search: [www.google.com](http://www.google.com); [www.yahoo.com](http://www.yahoo.com);  
[www.bing.com](http://www.bing.com); [www.ask.com](http://www.ask.com)
- Meta search: [www.dogpile.com](http://www.dogpile.com); [www.metacrawler.com](http://www.metacrawler.com);  
[www.clusty.com](http://www.clusty.com); <http://find.copernic.com>; [www.surfswax.com](http://www.surfswax.com);
- Directories: <http://dir.yahoo.com/>; [www.google.com/dirhp](http://www.google.com/dirhp);  
[www.questia.com/Index.jsp](http://www.questia.com/Index.jsp) [www.eldis.org/](http://www.eldis.org/);  
<http://openlearn.open.ac.uk/>; [www.dmoz.org](http://www.dmoz.org)
- Deep Web: [www.science.gov](http://www.science.gov);  
[www.deepwebtech.com/](http://www.deepwebtech.com/); <http://infomine.ucr.edu/>;  
[www.completeplanet.com/](http://www.completeplanet.com/); [www.stumbleupon.com/](http://www.stumbleupon.com/)





# Assignment

- When was the first CTC and where?
- Find 5 job vacancies in international organizations based in the Caribbean
- Find the top sites in health, entertainment, news and trade in the Caribbean
- Find photographs of the earliest Catholic missions to 3 Caribbean countries.

