

The Hidden...Invisible...Deep Web

What is the Hidden Web?

The Hidden Web (also known as the Invisible Web or the Deep Web) refers to that large part of the Internet that is not accessible through Google or other ordinary search tools.

Most of the Hidden Web, while presenting a possible rich vein of quality information, is only accessible through specialized databases or search engines.

The reason the information remains hidden is because most search engines cannot get past home pages to search within for relevant information.

Six reasons pages are invisible

The search engine chooses not to find some pages. They may be embedded in large sites that waste an engine's 'finding resources'. Some content may be found if it is linked to or promoted by others.

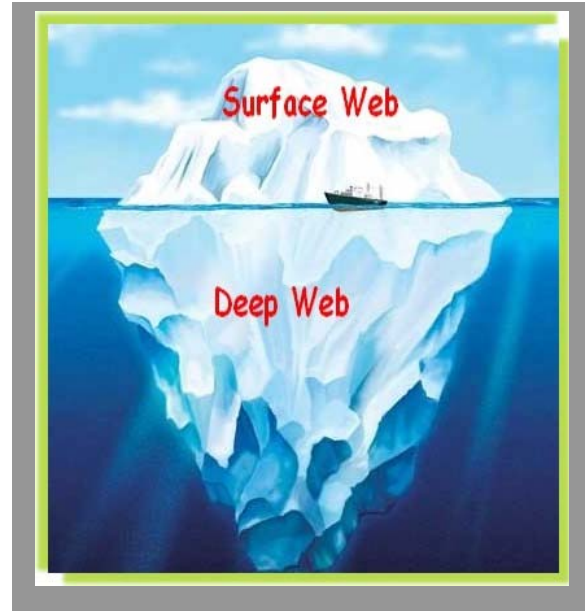
Disconnected pages. Search engines use crawlers that look for links from pages they already know. These pages have no one linking to them.

Non-indexed pages. Some pages have coding that excludes search robots.

Password-protected pages. Search engines will not index pages that it cannot enter.

Formatted content. A search engine looks for textual indicators to match keywords to pages. It is still impossible to do a visual or audible match without a text description.

Pages within a database. The search engine will not know what to put in a database search box, so ignores these sites. This includes dynamically-generated content like prices and statistics.



The good news is you can still access this material.

Portals: pathways created to relevant web pages and can provide personalized capabilities.

Gateways: web pages designed to attract search engines and visitors whose content may not be indexed by the search engines.

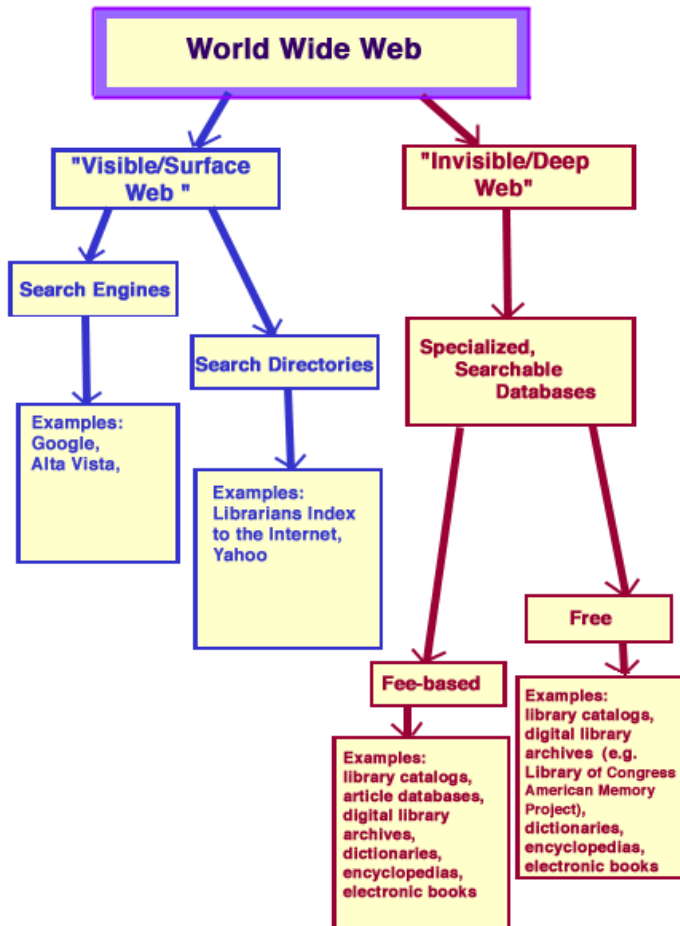
Directories: pre-selected sites compiled by a group of editors that are grouped under subject headings.

Databases: collections of documents with their own search engines. Can be password-protected or open for public access.

Resource guides: subject-specific web guides created by instructors and librarians



Tips and Tricks



Graphic representation of the World Wide Web and levels of Hidden Web content

"...d'oh (interjection)
Expressing frustration at the realization that things have turned out badly or not as planned.
(Oxford English Dictionary)..."

ACCESS to DATABASES

SUBSCRIPTION: KES

(username kesbath, password library)
 search.ebscohost.com, worldbookonline.com
 OLIVER catalogue

PUBLIC LIBRARY

(library card + pin number)
 OED, National Biography, Debrett's, NewsUK, Oxford Art/English/Music

TIPS FOR GOOGLING

How does it work? A web crawler goes through the page looking at page title, headings, page descriptors and follows links. It ranks pages highly that have the most or most longstanding links to it.

Getting the most from Google

Use the plus sign (+) to force a search for an overly common word. Use the minus sign (-) to exclude a term from a search. No space follows these signs. To search for a phrase, supply the phrase surrounded by double quotes (" ").

A period (.) serves as a single-character wildcard.

An asterisk (*) represents any word—not the completion of a word, as is traditionally used.

Use Google Scholar or Google Books

Use Advanced Search

define: for Google to find the definition of a term

The Google synonym operator is '~'

Use **site:** to limit your search to a site or site type

Use **filetype:** followed by a file extension but don't include a full stop before it. filetype:pdf

The **link:** instructs Google to look within hyperlinks for the search time

intitle: tells Google to search for a term within the title of a document

inurl: means only look in the web address for the selected term

related: finds pages related to a specific page

LIST of FREE DATABASES by SUBJECT AREA

Books Online

There are tens of thousands of classic books with full text available online for free.

[The Online Books Page](#) —
digital.library.upenn.edu/books/search.html

A searchable database of more than 25,000 English works with full text available for free online.

[Bibliomania](#) —
www.bibliomania.com/ A searchable database of free online literature from more than 2,000 classic texts.

[Project Gutenberg](#) —
www.gutenberg.org/ebooks/
A searchable online catalog of more than 19,000 free books with full text available online.

[The National Academies Press](#) —
www.nap.edu/browse.html A searchable directory of 3,000 free books online and 900 for-sale PDFs.

[ebrary](#) —*shop.ebrary.com/*
A searchable database of more than 20,000 full-text books. Small subscription fee.

Art

Hundreds of other museums all over the world have their own databases.

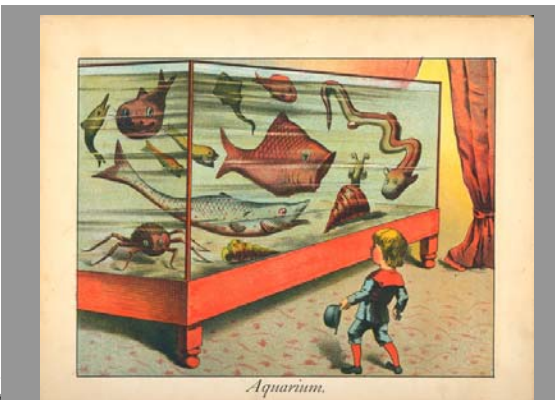
[Tate](#) —*www.tate.org.uk/* All the Tate museums in one searchable spot on the internet including a large range of British artists.

[Musée du Louvre](#) —*www.louvre.fr* One of the oldest and most famous art galleries in the world. Contains Leonardo da Vinci's *Mona Lisa* and Alexandros de Antioch's *Venus de Milo*.

[Guggenheim Museum](#) —
www.guggenheimcollection.org/ A collection of over 160 classical and new artists, searchable by artist name, title, date, movement, medium, concept, and museum.

[Smithsonian National Portrait Gallery](#) —
www.npgportraits.si.edu/ A searchable database of more than 80,000 portrait records.

[The National Gallery of Art](#) — *www.nga.gov/* A searchable catalog of all of the museum's 110,000 objects, with images of more than 6,000 available online.



Statistics

A range of sources of government and other information, mostly free but some commercial content that can be ignored.

[Official Statistics on the Web](#)
www.offstats.auckland.ac.nz

One of the best starting point for statistical sources by country, topic or subject. This service includes sources offering free and easily accessible social, economic and general data from official or similar "quotable" sources.

[The British Population Survey](#)
http://www.thebps.co.uk/

The British Population Survey interviews between 6,000 and 8,000 individuals every month (80,000 to 85,000 a year).

[National Statistics - the official UK statistics site](#)
www.statistics.gov.uk

[Social Trends](#) *www.statistics.gov.uk/socialtrends/*
Published by the UK National Statistics Office, Social Trends "draws together social and economic data from a wide range of government departments and other organisations to paint a broad picture of British society today, and how it has been changing.

[UNICEF Statistics](#) *www.unicef.org/statis/*
Statistics organised by country and giving population, GNP, fertility rate, immunisation levels, infant mortality, literacy rate.

[The World Gazetteer](#) *world-gazetteer.com*
Current population figures for cities, towns and places of all countries.

Science

[Highwire](#) *http://highwire.stanford.edu/*
2,099,134 free full-text articles as of 5/26/11
HighWire Press at Stanford University develops and maintains the Web versions of important journals in biomedicine and other disciplines.

[The NASA Astrophysics Data System](#)
http://adswww.harvard.edu/

-- approximately 300,000 free full-text articles
The NASA Astrophysics Data System (ADS) is a NASA-funded project that provides free access to the full text of articles in astronomy and astrophysics.

GENERAL HIDDEN WEB SEARCHING

[Complete Planet](http://aip.completeplanet.com/) - aip.completeplanet.com/ calls itself the 'front door to the Deep Web'. This free and well designed directory resource makes it easy to access the mass of dynamic databases that are cloaked from a general purpose search. The databases indexed by Complete Planet number around 70,000 and range from Agriculture to Weather. Also thrown in are databases like Food & Drink and Military.

[DeepWebTech](http://www.deepwebtech.com/) - www.deepwebtech.com/ gives you five search engines (and browser plugins) for specific topics. The search engines cover science, medicine, and business. Using these topic specific search engines, you can query the underlying databases in the Deep Web.

[IncyWincy](http://www.incywincy.com) - www.incywincy.com is an Invisible Web search engine. Acting as a meta-search engine it taps other search engines and filters the results. It searches the web, directory, forms, and images. With a free registration, you can track search results with alerts.

[Infomine](http://infomine.ucr.edu/) - infomine.ucr.edu/ has been built by a pool of libraries in the United States. Infomine 'mines' information from databases, electronic journals, electronic books, bulletin boards, mailing lists, online library card catalogs, articles, directories of researchers, and many other resources.

[Infoplease](http://www.infoplease.com/) - www.infoplease.com/ is an information portal with a host of features. Using the site, you can tap into a good number of encyclopedias, almanacs, an atlas, and biographies.

[Intute](http://www.intute.ac.uk/) - www.intute.ac.uk/ is UK centric and has some of the most esteemed universities of the region providing the resources for study and research. You can browse by subject or do a keyword search for academic topics like agriculture to veterinary medicine.

[Pinakes](http://www.hw.ac.uk/libwww/irm/pinakes/pinakes.html) - www.hw.ac.uk/libwww/irm/pinakes/pinakes.html a self-described launchpad into the internet. Arranged in subject areas, it provides links to some of the best academic sites on the web. Hosted by Heriot-Watt University in Edinburgh.

[Scirus](http://www.scirus.com) - www.scirus.com has a pure scientific focus. It is a far reaching research engine that can scour journals, scientists' homepages, courseware, pre-print server material, patents and institutional intranets.

[TechXtra](http://www.techxtra.ac.uk/) - www.techxtra.ac.uk/ concentrates on engineering, mathematics and computing. Feedback is from industry news, job announcements, technical reports, technical data, full text eprints, teaching and learning resources along with articles and relevant website information.

[Teoma](http://s.teoma.com/) - s.teoma.com/ Teoma is the search engine that powers Ask. The core of the Teoma search technology is based upon the idea that society and the Internet consist of tiny communities that self-organize themselves into hubs and authorities. Because of this ability to track users' searches, it aims to find concepts instead of keywords.

IMAGES

Alternative image searches offer the opportunity for historic, artistic and scientific imagery as well as digitized books.

[New York Public Library](http://digitalgallery.nypl.org/nypldigital/index.cfm)

<http://digitalgallery.nypl.org/nypldigital/index.cfm>
The Digital Gallery provides free and open access to over 700,000 images digitized from the The New York Public Library's vast collections, including illuminated manuscripts, historical maps, vintage posters, rare prints, photographs and more.

[The Courtauld Institute of Art](http://www.artandarchitecture.org.uk/index.html)

<http://www.artandarchitecture.org.uk/index.html>
Beautiful and inspiring images from the Courtauld.

[National Statistics - the official UK statistics site](http://www.statistics.gov.uk)

www.statistics.gov.uk
The WorldImages database provides access to the California State University IMAGE Project. It contains almost 75,000 images, is global in coverage and includes all areas of visual imagery.

[VADS](http://www.vads.ac.uk/) <http://www.vads.ac.uk/>

The UK's VADS (Visual Arts Data Service) is a consortial project with digital image collections in archaeology, history, the visual and performing arts, and literature/languages/linguistics.

CLASSICS

[Library of Ancient Texts Online](http://sites.google.com/site/ancienttexts/Home)

<http://sites.google.com/site/ancienttexts/Home>
Large alphabetized database of ancient texts.

[ABZU](http://www.etana.org/abzubib) <http://www.etana.org/abzubib>

Abzu is a guide to networked open access data relevant to the study and public presentation of the Ancient Near East and the Ancient Mediterranean world.

EARLY YEARS

[International Children's Digital Library](http://www.childrenslibrary.org/)

<http://www.childrenslibrary.org/>
Fully digitized children's books, both new and classic. Super!

[QuinturaKids](http://quinturakids.com/)

<http://quinturakids.com/>
large text and a graphic search availability

[KidRex](http://www.kidrex.org/)

<http://www.kidrex.org/>
If you're a kid who loves dinosaurs and you're looking for information on the web, this is your stop. Too bad about the Google ads©

[Factmonster](http://www.factmonster.com/)

www.factmonster.com/
loads of facts, highly engaging and colourful