

Deep Web Research

Lei Jin and Jay Wolofsky

Ryerson University Library

Deep Web/Surface Web

The Deep Web contains high quality structured information not accessible from conventional search engines such as Google.



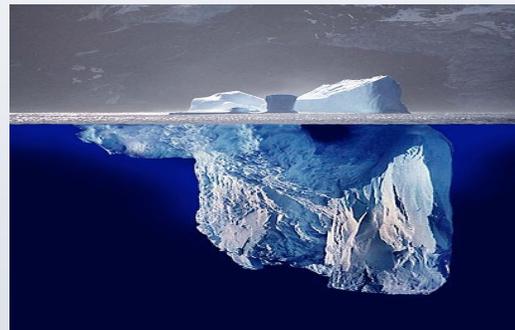
The spidering and crawling technology used by conventional search engines returns links based on popularity, not content.



Surface Web search results are ranked by the frequency documents link to each other (page rank). The first results are those that have had the most references by other documents, and not necessarily the most relevant or recent information.

The Deep Web

The Deep Web is several orders of magnitude larger than the surface in this parallel world (currently 400 to 500 times larger than the commonly defined Surface Web or 7,500 terabytes of information compared to 19 terabytes of information in the Surface Web).



Information in the Deep Web is buried far down on dynamically generated sites. Traditional search engines cannot find or retrieve information from the Deep Web because those pages don't exist until they are created as the result of a specific search.

Federated Search Engines



Federated search engines execute simultaneous real time searches of the Deep Web using very sophisticated software "connectors".

Accessing Deep Web Content



Biznar.com (Business)

Cepal.DeepWebAccess.com (Multidisciplinary)

CompletePlanet.com (Multidisciplinary)

DeepDyve.com (Multidisciplinary)

E-Print Network osti.gov/eprints/ (Science and Technology)

Highbeam.com (Multidisciplinary)

Infomine.ucr.edu (Multidisciplinary)

JGatePlus.com (Science and Technology)

Mednar.com (Medicine)

OpenDOAR.org (Academic and Institutional Depositories)

Science.gov (Science and Technology)

ScienceResearch.com (Science and Technology)

Scirus.com (Science and Technology)

SocialScienceResearchNetwork ssrn.com (Social Sciences)

WorldWideScience.org (Science and Technology)

(--THE

This Power
(version 20
commonly
If you are
template f

Verifying t
Go to the v
preferred m
the size of
be printed
poster will
100% and e
before you

Using the
To add text
placeholder
a placeholder
your cursor
to this sym
its new loc
Additional
side of this

Modifying
This templ
different c
Right-click
on the bac
click on "L
the layout
The colum
cannot be
layout by g

Importing
TEXT: Past
placeholder
left side of
needed.
PHOTOS: D
click in it a
TABLES: Yo
external do
adjust the
table that
click FORM
change the

Modifying
To change
the "Design
choose from
can create

© 2012 P
2117 For
Berkeley
E-Print
Technolo

Highbeam.