

## فراسوی گوگل: وب پنهان

## وب پنهان چیست؟

«وب پنهان» مفهومی است که برای اطلاعاتی که بر روی شبکه جهانی وب است، اما با استفاده از موتورهای جستجوی همه منظوره یافت نمی‌شود.

## واقعیتی درباره وب پنهان

- بهترین موتورهای جستجو فقط قادر به در دسترس قرار دادن حدود ۱۴٪ اطلاعاتی هستند که بر روی شبکه جهانی وب قرار داده شده است، بنابراین ۸۶٪ اطلاعات خارج از دسترس موتورهای جستجو است.
- اندازه وب پنهان ۵۵۰ بیلیون بزرگتر از وب آشکار ظاهر است.
- وب پنهان حاوی حدود ۵۵۰ بیلیون مدرک است، در حالیکه در مقام مقایسه، با وب آشکار یک بیلیون صفحه است.
- وب پنهان بزرگترین رشد مقوله اطلاعات جدید را در وب دارد.
- ۹۵٪ وب پنهان اطلاعات برای همگان قابل دسترسی است (اما از طریق موتورهای جستجو نمی‌توان به آنها رسید).
- بیش از نیمی از وب پنهان در پایگاه‌های اطلاعاتی معین است.



## ویژگیهای محتوایی وب پنهان: چرا موتورهای جستجو

## نمی‌توانند اطلاعات را بیابند

- محتوا در پایگاه اطلاعاتی ذخیره شده است
- پایگاه اطلاعاتی با پرداخت هزینه قابل دسترسی است
- محتوا در زمان واقعی ارائه می‌شود
- قالبهای محتوا
- ورود به سایتها، نیازمند اجازه ویژه است
- تعاملی بودن محتوا
- محتوا به عنوان نتیجه کاوش کاربر، بصورت پویا تولید می‌شود.
- سایت‌هایی که توسط پروتکل‌های بازدارنده رباتها مسدود شده‌اند

- اطلاعاتی که به تازگی به سایت افزوده شده و موتورهای جستجو هنوز به آنها دسترسی نیافته‌اند.

| نمونه‌ها  | رهای استاندارد یافت نشده است   |
|---|--|
| پایگاه‌هایی مانند:<br>ERIC, PubMed, Library catalogs    | لاعاتی (با درخواست کاوش کاربر، محتوا تولید می‌شود)   |
| قالب‌های پی.دی.اف، دیداری و شنیداری و هر قالب جدید دیگر | قالب‌های محتوا   |
| وضعیت آب و هو، قیمت سهام، وضعیت پرواز هواپیماها         | عات بلادرنگ (زمان واقعی)   |
| سایت‌های:<br>EBSOchost, INFOTRAC, LexisNexis Academic   | خدمات نیازمند پرداخت هزینه است   |
| همه سایت‌هایی که نیازمند عضویت هستند                    | سایت نیازمند رمز عبور یا ثبت نام است   |
| سایت‌های راهنمای مسافرت، سایت‌های کاریابی               | سایت نیازمند پر کردن فرم خاصی است  |
| سایت‌های خصوصی  | نمایه‌ای (درج تگ‌هایی که کاوشگرها را از نمایه ن آنها باز می‌دارد مانند تگ<br><META NAME="ROBOTS" CONTENT="NO |

## راه‌های که دسترسی به وب پنهان را هموار می‌کند:

### ۱. راهنماها:

- ▶ Open Directory Project: [www.dmoz.org](http://www.dmoz.org)

بزرگترین راهنمای وب که ویراستاری آن توسط افراد (غیر ماشینی) انجام می‌گیرد

- ▶ Librarian's Internet Index: [www.lii.org](http://www.lii.org)

پایگاه اطلاعاتی با حدود ۱۷۰۰۰ وب سایت با عناوین متنوع. با ارائه خبرنامه

- ▶ Infomine: <http://infomine.ucr.edu>

مجموعه پژوهش‌های اینترنتی دانش‌پژوهان از دانشگاه کالیفرنیا جنوبی

- ▶ Digital Librarian: [www.digital-librarian.com](http://www.digital-librarian.com)

سایت «کتابدار دیجیتالی» مارگارت اندرسون – کتابدار

### ۲. مسیرها

- ▶ Library and Information Technology Association:

<http://www.ala.org/ala/lita/litaresources/toolkitforexpert/toolkitexpert.htm>

انجمن کتابداری و فناوری اطلاعات

▶ A tutorial from the University of California-Berkeley:  
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html>

خودآموزي از دانشگاه کالیفرنیا-برکلی

▶ Lackie, Robert J. "Those Dark Hiding Places: The Invisible Web Revealed."  
[http://library.rider.edu/scholarly/rlackie/Invisible/Inv\\_Web\\_Main.html](http://library.rider.edu/scholarly/rlackie/Invisible/Inv_Web_Main.html)

آشکار سازی وب پنهان "آنانی که مخفیگاهها را تاریک می‌کنند"

## ۳. کاوشگرها

- ▶ BrightPlanet's [www.completeplanet.com](http://www.completeplanet.com)  
[www.kartoo.com](http://www.kartoo.com)
- ▶ مناسب برای یافتن فایل‌های قالب پی.دی.اف و گروه‌های خبری [www.google.com](http://www.google.com)
- ▶ [www.search-22.com](http://www.search-22.com)
- ▶ [www.freepint.com/gary/direct.htm](http://www.freepint.com/gary/direct.htm)
- ▶ <http://searchpdf.adobe.com>
- ▶ [www.closerlooksearch.com](http://www.closerlooksearch.com)
- ▶ دسترسی به مجموعه منابع کتابخانه دیجیتال <http://oaister.umd.umich.edu/o/oaister/>

## ۴. پایگاه‌های اطلاعاتی خاص

- ▶ [www.findarticles.com](http://www.findarticles.com) - a database of free articles
- ▶ [MagPortal.com](http://MagPortal.com) - a database of free articles
- ▶ [www.anywho.com](http://www.anywho.com) - "Finding People, Places, and Businesses"
- ▶ [www.findlaw.com](http://www.findlaw.com) - information on federal and state laws, lawsuits, etc.
- ▶ <http://www.thegateway.org> - Gateway to Educational Materials, U.S. Dept. of Education - cataloged collection of educational materials from federal, state, and university web sites.
- ▶ [www.nsdI.org](http://www.nsdI.org) - National Science Digital Library
- ▶ [www.usa.gov](http://www.usa.gov)
- ▶ [www.gpoaccess.gov](http://www.gpoaccess.gov)
- ▶ <http://catalog.loc.gov> - Library of Congress
- ▶ [www.ncbi.nlm.nih.gov/PubMed](http://www.ncbi.nlm.nih.gov/PubMed) - free Web version of Medline
- ▶ [www.infousa.com](http://www.infousa.com) - company database of 12 million records
- ▶ [www.singingfish.com](http://www.singingfish.com) - For audio and video searching
- ▶ <http://www.scienceresearch.com/> - Searches web-based science journals, databases, and websites not accessible through general purpose search engines
- ▶ [www.shadowtv.com](http://www.shadowtv.com)

## ۵. چگونه بهنگام باشیم

- ▶ <http://searchenginewatch.com>
- ▶ [www.marylaine.com/neatnew.html](http://www.marylaine.com/neatnew.html)
- ▶ [www.researchbuzz.com/wp](http://www.researchbuzz.com/wp)
- ▶ [www.lii.org/search/file/maillinglist/](http://www.lii.org/search/file/maillinglist/)
- ▶ [www.freepint.com/issues/issues.htm](http://www.freepint.com/issues/issues.htm) - newsletter
- ▶ Subscription to various listservs, discussion lists is one other way to keep up with the changing Web environment.
- ▶ [www.resourceshelf.com](http://www.resourceshelf.com)
- ▶ [www.docuticker.com](http://www.docuticker.com)

## ۶. وبلاگها

- ▶ Deep Web Research - <http://deepwebresearch.blogspot.com/>

An information Blog created and maintained by Marcus P. Zillman.

► Invisible Web Weblog - <http://invisibleweblog.blogspot.com/>

Maintained by Yazdan Mansourian, a research student at the University of Sheffield, England.

این مقاله از منبع زیر تلخیص شده است:

Jane, Devine & Egger-Sider, Francine. Beyond Google: The Invisible Web

<http://www.lagcc.cuny.edu/library/invisibleweb>

#### شرایط استفاده از مقاله

۱- انتشار ( چاپی یا الکترونیکی) مقالات بدون اخذ مجوز از ناشر(پژوهشگاه اطلاعات و مدارك علمی ایران) ممنوع است و تخلف از آن موجب پیگرد قانونی خواهد شد.

۲- فقط نقل چکیده مقالات در سایر محیطهای الکترونیکی (مانند سایتهای اینترنتی) مجاز است که باید همراه با ذکر نام نویسنده ، نام ناشر و درج نشانی پیوند(لینک) اینترنتی پژوهشگاه باشد.