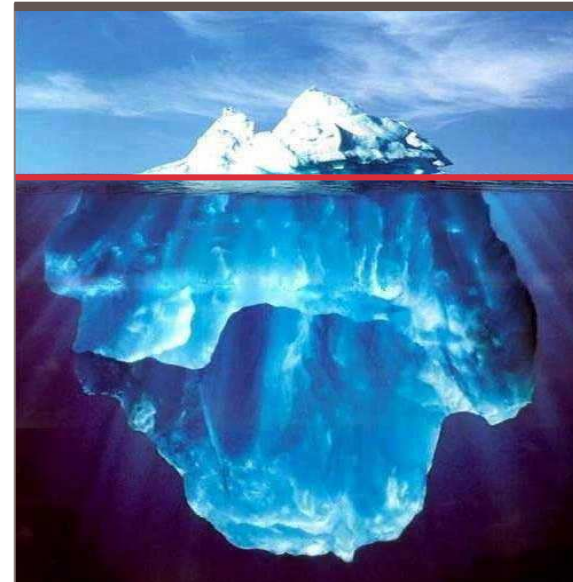


Common barriers to web search engine crawling

Barriers to web search engine crawling

What can make a site effectively invisible to search engine users

- Content “hidden” behind search forms
- Non-HTML links
- Outdated robots.txt crawling restrictions
- Server errors (crawler times out when fetching content)
- Orphaned URLs
- Rich media: audio, video
- Premium content



WEB
Searchable

DEEP WEB
Not searchable

Purpose of illustrations

- The following illustrations identify barriers to web search engine crawling common to government websites
- Such barriers are found on the majority of federal agency and inter-agency sites
- Illustrations:
 - Business professional seeking export opportunities
 - Local historian researching World War II
 - Entrepreneur seeking ATF's program for small businesses
 - Representative federal agency's sitemapping target list

Business professional seeking export opportunities

- Sample search on Google Web Search: [shenyang china retail outlet stores]

(<http://www.google.com/search?hl=en&safe=off&q=shenyang+china+retail+outlet+stores>)

- What a search engine user does not find:

http://www.export.gov/eac/show_short_trade_leads.asp?Country=156&Submit2=View+Leads

- Reason: Drop-down menu/underlying database renders Export.gov/US Commercial Service Trade Leads database uncrawlable: http://www.export.gov/eac/trade_leads.asp

[Personalize this page](#) | [Sign in](#)



[Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [Desktop](#) [Moma](#) [more »](#)

shenyang china retail outlet stores

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

Google Search I'm Feeling Lucky

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2007 Google



Web Images Video News Maps Desktop Moma more »
shenyang china retail outlet stores Search Advanced Search Preferences

Web Results 1 - 10 of about 24,100 for shenyang china retail outlet stores. (0.16 seconds)

[Dangdang.com Plans To Open More Retail Outlets - ChinaTechNews.com ...](#)

Dangdang.com Plans To Open More **Retail Outlets** - ChinaTechNews.com. ... **Shenyang** To Host **China's** Largest Automatic Identification Manufacturing Park ...
www.chinatechnews.com/2005/12/12/3187-dangdangcom-plans-to-open-more-retail-outlets/ - 59k - [Cached](#) - [Similar pages](#)

[Nokia Opens First Flagship Store In Asia - ChinaTechNews.com - The ...](#)

"Nokia is committed to pushing the boundaries of the **retail** landscape." ... **Shenyang** To Host **China's** Largest Automatic Identification Manufacturing Park ...
www.chinatechnews.com/2006/07/17/4236-nokia-opens-first-flagship-store-in-asia/ - 58k - [Cached](#) - [Similar pages](#)

[\[PDF\] Investing in China's Retail Industry](#)

File Format: PDF/Adobe Acrobat - [View as HTML](#)
The number of **outlets** controlled by **China's** top 30 domestic chain **stores** ... first entered the DIY/home improvements market in Beijing, then **Shenyang**. ...
[www.pwc.com/.../pwcpublications.nsf/docid/4B47E208388DCF208525718E006EE2E2/\\$FILE/retail-china-april-2006.pdf](http://www.pwc.com/.../pwcpublications.nsf/docid/4B47E208388DCF208525718E006EE2E2/$FILE/retail-china-april-2006.pdf) - [Similar pages](#)

[\[PDF\] 4 2004/2005 Global Retail & Consumer Study from Beijing to ...](#)

File Format: PDF/Adobe Acrobat - [View as HTML](#)
formats and the increase in **retail outlets**, the fast. moving consumer goods market has been growing ... across major cities in **China**. Typically these **stores** ...
www.pwc.com/gx/eng/about/ind/retail/growth/china18-31pgs.pdf - [Similar pages](#)

[Retailers, Supermarkets, Chain Stores, Retail Chains China Asia](#)

Sponsored Links

[China Outlet Stores](#)

Bargain Prices. Smart Deals. Save on **China Outlet Stores!**
[Shopzilla.com](#)

[China Outlet Stores](#)

Find Deals on **China Outlet Stores** "Compare Prices, Brands, and More"
[www.smarter.com](#)

[Discount China](#)

Find great offers for pottery & **china** and save up to 75% now!
[www.best-price.com/Pottery&Glass](#)

[Discount China](#)

Looking to find discount **China**? Visit our discount **China** guide.
[Discount-China.info](#)

[China Outlet Stores](#)

Looking for **china outlet stores**? Save!
[www.shoppingpage.org](#)



- WCMP 2-15-07.ppt
- Examples of barriers ...
- atf - Google Search - ...
- USASearch.gov: atf - ...
- shenyang china retail...
- enemy prisoner of wa...

Internet 100%
1:11 AM
Thursday
2/15/2007

EXPORT.GOV

Helping U.S. Companies Export

Register | Login

Search Export.gov

Advance Search

Home | About Export.gov | Partner Agencies | Contact Us | Non-U.S. Companies

Print | E-mail Page

Find Opportunities

- ▶ By Industry
- ▶ Market Research
- ▶ Trade Events
- ▶ Trade Leads

Find Solutions

- ▶ International Sales-Marketing
- ▶ International Finance
- ▶ International Logistics
- ▶ Regulations & Licenses
- ▶ Trade Data & Analysis
- ▶ Trade Problems

Contact Us

1-800-USA Trade

- ▶ Find a Local U.S. Office
- ▶ Find an Overseas Office

Trade Lead Information By Country For China

Looking for wire bonding capillary and pick up tool manufacturers

Location: Shanghai, China Expiration: 06/30/2007 Published: 11/07/2006

Details: Shanghai HaoKe Microelectronics Company is a specialized distributor in th

Shenyang Retail Group Seeks U.S. Designer Outlet Stores

Location: Shenyang, China Expiration: 06/30/2007 Published: 11/06/2006

Details: The Shenyang Commercial City Group, owned by the government of Shenyang ci

Wants to purchase 200T polyethylene per month and looking for agent opportunity of raw materials of blow moulding products

Location: Wuhan, China Expiration: 09/25/2007 Published: 10/11/2006

Details: Wants to purchase 200T polyethylene per month and looking for agent oport

Looking for Septum and agent opportunity of lithium Batteries

Location: Wuhan, China Expiration: 09/25/2007 Published: 10/11/2006

Details: Looking for Septum and agent opportunity of lithium Batteries

Cold Rolling Stainless Steel Board

Location: Wuhan, China Expiration: 09/25/2007 Published: 10/11/2006

Details: The company wants to buy the Cold Rolling Stainless Steel Board produced b

Looking for chemicals for construction applications

Location: Shanghai, China Expiration: 06/30/2007 Published: 07/04/2006

Details: Megachem International Trading Company is headquartered in Singapore. Mega

start

Examples of barriers ...

show_short_trade_le...

Trade Leads - Windo...

Internet

100%

9:32 PM
Wednesday
2/14/2007

EXPORT.GOV

Helping U.S. Companies Export

Register | Login

Search Export.gov

Advance Search

Home | About Export.gov | Partner Agencies | Contact Us | Non-U.S. Companies

Print | E-mail Page

Find Opportunities

- ▶ By Industry
- ▶ Market Research
- ▶ Trade Events
- ▶ Trade Leads

Find Solutions

- ▶ International Sales-Marketing
- ▶ International Finance
- ▶ International Logistics
- ▶ Regulations & Licenses
- ▶ Trade Data & Analysis
- ▶ Trade Problems

Contact Us

1-800-USA Trade

- ▶ Find a Local U.S. Office
- ▶ Find an Overseas Office



Search Trade Leads

The following are new, pre-qualified trade opportunities and foreign government tender announcements that are available to U.S. exporters.
Choose either **country and/or industry**:

Search By Keyword:

enter a phrase

Search

OR

- Chad
- Chile
- China
- Colombia
- Congo (Democratic Republic)

Search By Country:

(Press and hold "Ctrl" button for **multiple selections**)

Search By Region:

=== Select Region =====
Africa

Local historian researching World War II

- Sample search on Yahoo Search: [enemy prisoner of war information bureau utah]

(http://search.yahoo.com/search;_ylt=A0geup1XltRF5SsAbs5XNyoA?p=enemy+prisoner+of+war+information+bureau+utah&ei=UTF-8&fr=yfp-t-501&x=wrt)

- What a search engine user does not find:

http://arcweb.archives.gov/arc/arch_results_detail.jsp?&pg=1&si=0&nh=1&st=b

- Reason: NARA Archival Research Catalog (ARC) database uncrawlable: http://arcweb.archives.gov/arc/basic_search.jsp

1. [Spanish-American War - Wikipedia, the free encyclopedia](#)
... to Puerto Rico on a reconnaissance mission, sponsored by the Army's **Bureau of ...** The war provided both sides a common **enemy** for the first time since the end of ...
en.wikipedia.org/wiki/Spanish-American_War - 94k - [Cached](#) - [More from this site](#)
2. [American Civil War - Wikipedia](#)
Hyperlinked overview of the four-year conflict between the Federal government and the breakaway Confederate States of America. Includes sections on the causes of the
Category: [U.S. Civil War \(1861-1865\)](#)
en.wikipedia.org/wiki/American_Civil_War - 253k - [Cached](#) - [More from this site](#)
3. [US National Archives WW II POW Records](#)
... **Prisoners of War (World War II)**, Prepared by the **Prisoner of ...** Prepared by the **Enemy prisoner of War information bureau**, Ft. Holabird, Baltimore 19, MD.
www.gentracer.com/nara.html - 8k - [Cached](#) - [More from this site](#)
4. [Protected Persons](#)
... reported to the Army **Information Bureau** as "escaped and not captured. ... laws governing **prisoners of war**, but which is unsure which **enemy** combatants meet ...
lawofwar.org/Protected_Persons.htm - 177k - [Cached](#) - [More from this site](#)
5. [The New Yorker : talk : content](#)
... fully involved in the atrocities, which began with the execution of **prisoners**. ... increasingly unpopular **war** against an entrenched guerrilla **enemy**—to publish ...
newyorker.com/talk/content/?031110ta_talk_hersh - 56k - [Cached](#) - [More from this site](#)

ARC Search Results: Hit #1 of 1

1 hit(s) retrieved for keywords "logan utah war" .

[See Hierarchy](#) [Printer-Friendly Version](#) [Back to Search Results](#)
[Refine Search](#) [New Search](#)

Jump to Hit [Go](#)

ARC Identifier: 899706
Title: Camps Inactivated: 9th Service Command: [Camp] Logan, Utah
Creator: [War Department, Office of the Provost Marshal General, Enemy Prisoner of War Information Bureau](#), (ca. 1942 - 09/18/1947) (Most Recent)
Type of Archival Materials: Textual Records
Level of Description: File Unit from Record Group 389: Records of the Office of the Provost Marshal General, 1920 - 1975
Location: Modern Military Records LICON, Textual Archives Services Division (NWCTM), National Archives at College Park, 8601 Adelphi Road, College Park, MD 20740-6001 PHONE: 301-837-3510, FAX: 301-837-1752, EMAIL: inquire@nara.gov
Part of: [Series: Subject Files, 1942 - 1946](#)
Access Restrictions: Unrestricted
Use Restrictions: Unrestricted

Copy 1

Copy Status: Description, Reproduction, Reference

ARC Basic Search

Search for [descriptions](#) of [Archival Materials](#) containing the following keyword(s):

[Search Tips](#)

Go Clear
Advanced Search

Limit Results to 100 1000 2000

Filter Your Search:

[Descriptions](#) of [Archival Materials](#) linked to [digital copies](#)

[Archival materials dated](#) between MM DD YYYY and MM DD YYYY

[Location of Archival Materials:](#)

- All Locations
- Cartographic and Architectural Records LICON, Special Media Archives S...
- Center for Legislative Archives
- Civilian Records LICON, Textual Archives Services Division

[Type of Archival Materials:](#)

- All Types
- Architectural and Engineering Drawings
- Artifacts
- Data Files

Entrepreneur seeking ATF's program for small businesses

- Sample search on Microsoft Live Search: [atf small business program]
(<http://search.live.com/results.aspx?q=atf+small+business+program&mkt=en-US&form=QBRE>)
 - What a search engine user does not find:
<http://www.atf.gov/acquisition/smallbusiness.htm>
 - Reason: Robots.txt restriction renders ATF site uncrawlable:
<http://www.atf.gov/robots.txt>
- Lack of sitemap/web search management also accounts for outdated link to ATF's previous identity as Department of the Treasury agency in USA.gov (MSN) search results:
<http://usasearch.gov/search?v%3Aproject=firstgov-web&query=atf>

[Small Arms and Light Weapons: U.S. Policy and Views" - U.S. Foreign ...](#)
 TRACING ILLEGAL **SMALL ARMS: AN ATF PROGRAM** ... repository for all records of federal firearms licensees who have discontinued **business**.
usinfo.state.gov/journals/itps/0601/ijpe/pj62atf.htm · [Cached page](#)

[Business.gov: Business Development](#)
ATF's Small Business Program is the focal point of contact within the organization to provide advice and guidance on matters relating to the socio-economic programs.
www.sba.gov/businessgateway/business_development.html · [Cached page](#)

[Department of State Washington File: Byliner: ATF Official on Tracing ...](#)
 Tracing Illegal **Small Arms: An ATF Program** By Jacqueline K. Holmes (The author is a **program** manager ... repository for all records of Federal firearms licensees who have discontinued **business**.
usembassy-australia.state.gov/hyper/2001/0613/epf306.htm · [Cached page](#)

[Access to Finance](#)
 The Access to finance (**ATF**) **program** aims to boost MSME performance and ... partners interested in establishing a specialized Microfinance/**Small Business** ...
www.msme-nigeria.org/Actfin.htm · [Cached page](#)

SPONSORED SITES

[Dell Official Site](#)
 Get reliable technology customized to your specific **business** at Dell.
www.dell.com/smallbusiness

[Capital One Business Card](#)
 We think big about your **small business**. Apply online today!
capitalone.com/smallbusiness

[Small Business Class](#)
 Learn **Business** Plan Writing & More Visit Online for a Free Career Kit
pcdi.com

[Small Business](#)
 Run Your Company Remotely. Free Microsoft Mobile Biz Resource Kit!
www.microsoft.com/server

[Small Business Network](#)
 Largest network of **Small** Businesses find **small business** investors \$25k



Bureau of Alcohol, Tobacco, Firearms and Explosives

Acquisition and Business Opportunities

- What's New
- ATF Buyers
- What Does ATF Buy?
- ATF Contracting Opportunities
- ATF Small Business Program
- Federal Regulations
- ATF Internet
- Home

ATF Small Business Program

Anne D. Hudson, Chief Acquisition Management Branch
(202) 927-8200

Updated: 10/26/06

ATF's Small Business Program is the focal point of contact within the organization to provide advice and guidance on matters relating to the socio-economic programs. ATF maintains a positive role in participating in vendor outreach sessions.

Small Business Services

- Counseling available to small business concerns;
- Conduct capability briefings for customers seeking opportunities with small businesses;
- Link prime contractors to small businesses for subcontracting opportunities;
- Organize small business workshops and seminars; and
- Provide advice in small business issues.

Useful Links

SBA Ombudsman - www.sba.gov/ombudsman
SBA Programs - www.sba.gov/aboutsba/sbaprograms.html
Central Contractor Registration (CCR) www.ccr.gov

Google Desktop
Google Desktop will resume indexing in 60 seconds

```
User-agent: *  
Disallow: /
```


- All Results (112)
- Firearms, Alcohol, Tobacco (29)
 - Treas (27)
 - Sponsored (2)
 - DOJ, Oig Audit Report 06 (2)
 - Other Topics (52)

- ATF Office Locator - USA.gov
- Responsibilities of the ATF - USA.gov

112 results for atf out of at least 36,433 (Details)

Web results by msn Search

ATF Online [new window] [preview]
 The ATF is a law enforcement organization within the United States Department of Treasury. The BATF enforces the Federal laws and regulations relating to alcohol, tobacco, firearms, explosives and ...
www.atf.treas.gov More from Treasury Department

www.atf.gov [new window] [preview]
www.atf.gov More from Bureau of Alcohol, Tobacco, Firearms, and Explosives (Justice)

www.atf.treas.gov [new window] [preview]
www.atf.treas.gov/index.htm More from Treasury Department

ATF Online - Public Use Forms - Regulatory Enforcement Operations ... [new window] [preview]
 Problem viewing forms? ATF F 4473 PT. I (5300.9) Firearms Transaction Record Over-The-Counter - can not be posted, must be ordered from the Distribution Center
www.atf.treas.gov/forms/5000.htm - Cached - More from Treasury Department

Representative federal agency's sitemapping target list

- Sample list of sites with uncrawlable elements:

<http://spreadsheets.google.com/pub?key=pUb62ZKHnzggqEoGF4LFf3Gw>

EPA sites to sitemap: Sitemap

| Agency | Site/database name/description | Issue | URL |
|------------|--|---|---|
| | | Uncrawable--to sitemap | |
| | | Blocked by robots.txt | |
| | | Crawable | |
| EPA | EPA Online Library System (OLS) | Blocked by robots.txt; uncrawable--to sitemap | http://www.epa.gov/natlibra/ols.htm |
| EPA | Enforcement and Compliance database (enforcement penalties) | Uncrawable--to sitemap | http://www.epa.gov/echo/compliance_repo |
| EPA | Enforcement Case Reports Database | Uncrawable--to sitemap | http://www.epa.gov/echo/compliance_repo |
| EPA | Compliance database (Hazardous Waste section) | Uncrawable--to sitemap | http://www.epa.gov/echo/compliance_repo |
| EPA | EPA: Beach Watch: http://oaspub.epa.gov/beacon/beacon_national_page_main | Uncrawable--to sitemap | http://oaspub.epa.gov/beacon/beacon_nati |
| EPA | Environmental Information Mangement System (Document sub category in the advanced search) | Uncrawable--to sitemap | http://oaspub.epa.gov/eims/query.page |
| EPA | EPA: RadNet Database (Radiation levels) | Uncrawable--to sitemap | http://oaspub.epa.gov/enviro/erams_query |
| EPA | Database of single chemical toxicity for plant and animal species | Blocked by robots.txt; uncrawable--to sitemap | http://cfpub.epa.gov/ecotox/quick_query.ht |
| EPA | EPA Sponsored Employer Survey: Database of most commuter friendly work places | Uncrawable--to sitemap | http://www.ergweb.com/projects/cccli/search |
| EPA: NCER | National Center for Environmental Research http://es.epa.gov/ncer/ | Blocked by robots.txt | http://cfpub.epa.gov/ncer_abstracts/index |
| EPA: HPVIS | EPA High Production Volume Inormation System (HPVIS) http://www.epa.gov/hpvis/index.html | Uncrawable--to sitemap | http://iaspub.epa.gov/opthpv/public_search |
| EPA: GICS | Grants Information and Control System database | Uncrawable--to sitemap | http://www.epa.gov/enviro/html/gics/gics_c |
| EPA | Envirofacts Data Warehouse Land http://www.epa.gov/enviro/index.html | Uncrawable--to sitemap | http://oaspub.epa.gov/enviro/ef_home2.lan |