

Search Update

Webmasters User Group

BY KEVIN PADDOCK, DTS SEARCH
ADMINISTRATOR
"SEARCHMASTER@DTS.CA.GOV"

State of California Webmasters User Group

Friday, September 28, 2007

Search Update

2

Webmasters User Group



BY KEVIN PADDOCK, DTS SEARCH
ADMINISTRATOR
"SEARCHMASTER@DTS.CA.GOV"

State of California Webmasters User Group

Wednesday, April 9, 2008

Search Update - Agenda

3

1. What's New?
2. Free Google Site Search from DTS
3. Free Google Site Search from Google
4. DTS & eServices – Committed to Search
5. Search School – Cached Documents
6. Questions
7. T-Shirt!

State of CA Advertising Campaign

4



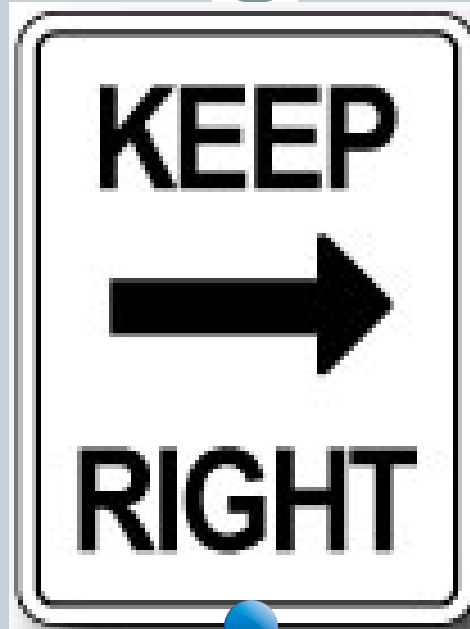
State of CA Advertising Campaign

5



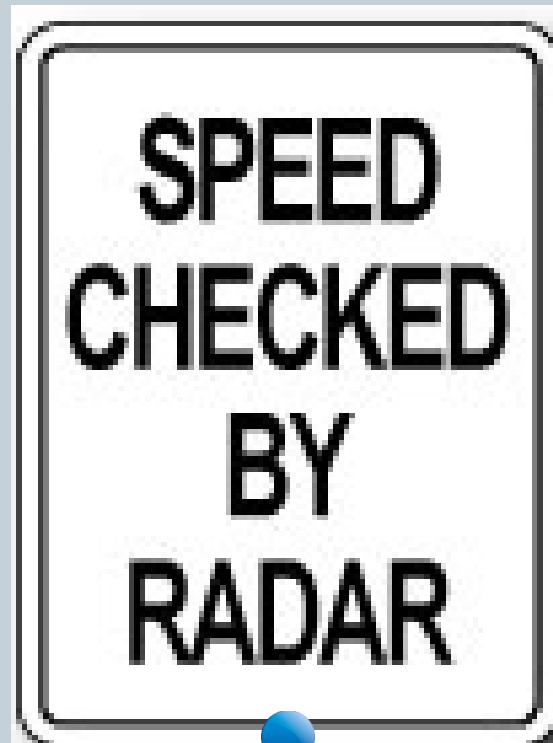
State of CA Advertising Campaign

6



State of CA Advertising Campaign

7



O'REILLY®

State of CA Advertising Campaign

8



Dewey, Cheetham & Howe
Attorneys at Law

1. What's New?

9

- Document license upgraded to 10 million docs
- Upgraded to GB-5005 clustered appliances
- More DTS staff trained to assist
- ASP SERP script now available
- Open Parameter PDF “#search”
 - Supported in Acrobat and Reader 7.0+
 - Supported in IE and Mozilla browsers

2. Free Google Site Search from DTS

10

- “DTS Site Search Service”
 - It’s free, so you won’t see it on the DTS Rate Schedule!
 - (did I mention it’s FREE?)
 - Many State web sites still have
 - Broken search
 - Poor search
 - No search
 - How about FREE Google search?
- To Get Started, read all about it at [CA Webtools](#)
- 140 (and counting) State of CA web sites use this service

3. Free Site Search from Google

11

- “Google Custom Search”
- <http://www.google.com/coop/cse/>
- Example: Consumer Services Center
 - <http://www.consumer-sc.ca.gov/>
- No ads for government sites

4. DTS & eServices – Committed to Search

12

- License upgrade from 2M to 10M documents
- Upgrade completed Feb. 29, 2008
- DTS Content Services Staff added to Search Team
- Collaboration with county, city and local municipalities in the works...
- Custom solutions using Google search.
- SERP scripts for more flexibility and control than XSLT-based results pages. So far we have:
 - [ASP – DTS Searchmaster](#)
 - PHP – DTS Searchmaster
 - JSP – volunteer?
 - .Net – volunteer?
- Want a GSA Admin Account: just call 916-464-4311

5. Search School – Cached Documents

13

- Which documents are cached on the GSA?
 - HTML, PDF, DOC
- Why do search engines cache documents?
 - Site down? Search Engine can display the pages
 - Client “reader” not required for application docs
 - Faster content delivery (since it’s HTML)
 - Faster query term discovery (search term highlighting)
 - Makes PDFs “Accessible”
- How Do I Access the Cached Version?
 - PDF, Doc: click “Text Version” in results
 - HTML: click “Cached”

Search School – Cached Documents

14

- Search CA for “strategic plan”

Search School – Cached Documents

15

- **How are docs converted to HTML?**
 - The GSA runs the open source PDFtoHTML converter
 - <http://pdftohtml.sourceforge.net/>
- **When cache (not cash) is a bad thing to have**
 - You remove unwanted files from your web site...
 - ...but the cached version still shows up in search engines

Search School – Cached Documents

16

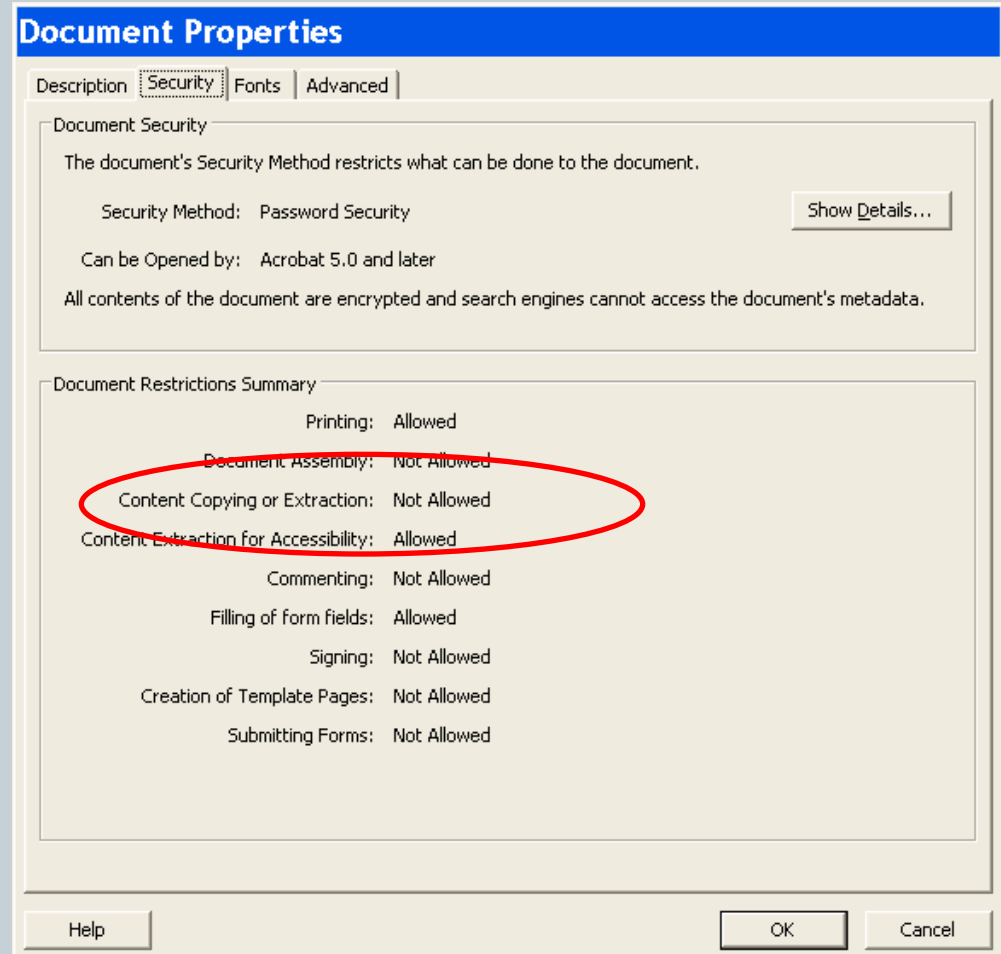
- To prohibit PDF caching:
 - Use Acrobat security
 - ✦ Webmaster or Document Provider responsibility
 - ✦ Prohibits caching on all search engines
 - Does not allow text copying.
 - The PDFtoHTML process cannot parse the document.
 - ✦ Example of doc with PDF Text Security:
 - [Search DMV for “address”](#)

Search School – Cached Documents

17

Properties View in Acrobat Reader 7.0

This document will not be cached on any search engine.



Search School – Cached Documents

18

Acrobat 7.0 Standard:
Document Properties
>>Security>>Password
Security

Permissions settings to
prohibit search engine
caching.

Password Security - Settings

Compatibility: Acrobat 5.0 and later

Encryption Level: High (128-bit RC4)

Select Document Components to Encrypt

- Encrypt all document contents
- Encrypt all document contents except metadata (Acrobat 6 and later compatible)
- Encrypt only file attachments (Acrobat 7 and later compatible)

All contents of the document will be encrypted and search engines will not be able to access the document's metadata.

Require a password to open the document

Document Open Password:

No password will be required to open this document.

Permissions

Restrict editing and printing of the document. A password will be required in order to change these permission settings.

Change Permissions Password:

Printing Allowed: High Resolution

Changes Allowed: None

Enable copying of text, images, and other content

Enable text access for screen reader devices for the visually impaired

Help OK Cancel

Search School – Cached Documents

19

- To prohibit HTML caching:
 - Use the Googlebot tag:
 - ✦ `<meta name="googlebot" content="noarchive">`
- How to Remove Cached Pages
 - From DTS GSA: call 464-4311, ask for “Remove URL from DTS search engine”
 - ✦ Indicate URL or URL pattern.
 - ✦ Blocking takes effect immediately.
 - ✦ DTS will also add the URL(s) to the “Do Not Crawl” list (which stops all further attempts to crawl the URL)
 - From Google Web: [Webmaster Tools](#)

Preguntas

21



Google T-Shirt “Contest”

22

- **Solve Puzzle to Win**
 - Random drawing from all correct answers
 - Send your answer to:
 - ✦ [searchmaster\[at\]dts.ca.gov](mailto:searchmaster@dts.ca.gov)
 - Deadline to submit your answer:
 - ✦ Friday April 18, 5:00 p.m.
 - It's not as hard as it looks...

Google T-Shirt “Contest”

23



X

4



Replace each Egyptian symbol with an Arabic numeral to make this multiplication equation correct. Each symbol always represents the same numeral.

Thanks for Watching

24

- **Contact:**
 - Kevin Paddock, Search Administrator
 - DTS: 464-4233

searchmaster@dts.ca.gov