
Webmasters User Group

September 20, 2006

Agenda

Introduction

California Portal Redesign Project Update

Look-and-Feel Update

Long-Term Strategy Update

Webmaster Toolbox

Google Search Engine

Questions and Answers, Discussion

Closing

Look-and-Feel Update

- Dane Wilson



Most Visited

1. [DMV](#)
2. [EVENTS IN CALIFORNIA](#)
3. [JOBS](#)
4. [STATE PHONE DIRECTORY](#)
5. [LEGISLATION](#)
6. [LICENSES](#)
7. [STATE LOTTERY](#)
8. [TAXES](#)
9. [UNCLAIMED PROPERTY](#)
0. [UNEMPLOYMENT](#)



Visit the [Governor's website](#)

The Governor saves millions for California.

- [Important item to see. Click somewhere to read something about](#)
- [Important item to see. Click somewhere to read something about](#)
- [Important item to see. Click somewhere to read something about](#)
- [Important item to see. Click somewhere to read something about](#)

What's New

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris justo libero, sodales a, pulvinar quis, aliquet vel, mauris. In aliquet massa a neque. Vestibulum aliquam nibh ut lorem. Duis facilisis, ante non molestie facilisis, urna diam tempor orci, in venenatis nunc mi nec purus. Praesent fringilla mollis tortor. Ut tempor aliquet lorem. In enim. Nulla enim. In hac habitasse platea dictumst. Donec convallis quam at urna sodales ultricies. Suspendisse quis eros.

Header Title

What's New

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris justo libero, sodales a, pulvinar quis, aliquet vel, mauris. In aliquet massa a neque. Vestibulum aliquam nibh ut lorem. Duis facilisis, ante non molestie facilisis, urna diam tempor orci, in venenatis nunc mi nec purus. Praesent fringilla mollis tortor. Ut tempor aliquet lorem. In enim. Nulla enim. In hac habitasse platea dictumst. Donec convallis quam at urna sodales ultricies. Suspendisse quis eros.

Aliquam at leo sit amet mi luctus ultrices. Etiam vehicula ante id tellus. Cras ullamcorper elit placerat metus pretium fringilla. Nunc congue. Fusce sem turpis, ultricies vel, porttitor vitae, faucibus sit amet, quam. Etiam at risus et ante posuere sollicitudin. Nam sapien lacus, accumsan a, molestie eu, tempus eu, sapien. Ut augue. Vivamus eget magna id nisl bibendum dapibus. Praesent mollis augue non ipsum.

Header Title



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris justo libero, sodales a, pulvinar quis, aliquet vel, mauris. In aliquet massa a neque. nar quis.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris justo libero, sodales a, pulvinar quis.

Webmaster Toolbox

- Megan Johnson, DTS
- Savita Farooqui, SymSoft Solutions





Online Webmaster Resource Guide and Training Manual

Savita Farooqui

savita@symsoftsolutions.com

The Team

- ▶ Skilled in Website Design and Development with focus on Usability, Accessibility and using Industry standards such as W3C
- ▶ Experience with creating Webmaster Resource Guides for many states

The Project: Breakdown

- ▶ **Online Webmaster Resource Guide**
 - Statewide standards for usability, accessibility and separation of content and display
 - Development tips for webmasters to plan, build, and maintain their web sites
 - Information about general web design principles and design tips
- ▶ **Webmaster Training Manual (Self-paced)**
 - Supports the principles of the Online Webmaster Resource Guide in a offline linear reading experience
 - Task oriented format addressing the needs for various skill sets
 - Readers learn at their own pace or use the manual in conjunction with a training curriculum
- ▶ **Proposed Webmaster Training Curriculum**
 - Sample topics:
 - Usability basics and testing
 - Effective web and application design
 - User Analysis and Conceptual Design

Current Status

- ▶ Project kickoff – week of 9/11/2006.
- ▶ Researching the CA State Standards, and current Webmaster Resource Guide
- ▶ Participated in IOUCA meetings and had initial discussions with IOUCA working groups
- ▶ Seeking Webmasters input through surveys; currently looking through received feedback and hope to get additional inputs
- ▶ Currently in the process of identifying components for the Webmaster User Guide and Training Manual

Next Steps

- ▶ Information organization for the Website - Create a draft outline for the Webmaster Resource Guide
- ▶ Gap analysis between current Webmaster Resource Guide and the proposed Webmaster Resource Guide
- ▶ Develop the Online Resource Guide
- ▶ Develop Training Manual
- ▶ Develop proposed Training Curriculum

Thank You

- ▶ Questions or more information?

Contact:

Megan Johnson

DTS Project Manager

Megan.Johnson@dts.ca.gov

Savita Farooqui

SymSoft Solutions Project Manager

savita@symsoftsolutions.com

Search Best Practices for Webmasters & Search System Update

by Kevin Paddock, DTS Search Administrator
“searchMaster@dts.ca.gov”

State of California Webmasters User Group

Wednesday, September 20, 2006



Current Search Engine

- Autonomy (formerly Verity) K2
- Currently provides search service for
 - Portal - www.ca.gov
 - DMV - www.dmv.ca.gov
 - Tourism - www.gocalif.ca.gov
 - Calohi - www.calohi.ca.gov
 - New customers in line for site search on GSA:
 - DMV, DHS, CTTC (tourism's new site), MRMIB, YOU?!

Search Statistics

- November of 2003, Portal searches:
 - 30,000 /mo.
- August 2006, Portal searches:
 - 944,936 /mo. (over 30,000 per day) 3000%
 - 277,621 unique queries
 - ~1M documents from 300 State web sites
 - (heavily filtered, lots more than 1M docs out there)
- August 2006 for DMV:
 - 722,642 searches (~23k per day)



Search Statistics

■ Top 20 Searches:

- | | |
|--------------------------------|-------------------------------|
| 1. dmv : 19760 | 11. employment : 3690 |
| 2. secretary of state : 12968 | 12. unemployment : 3326 |
| 3. forms : 9774 | 13. child support : 3243 |
| 4. unclaimed property : 5847 | 14. state parks : 3174 |
| 5. jobs : 5184 | 15. map : 3152 |
| 6. title 22 : 4317 | 16. medi-cal : 3076 |
| 7. birth certificate : 4311 | 17. birth certificates : 3045 |
| 8. edd : 4260 | 18. megans law : 3012 |
| 9. real estate : 4245 | 19. disability : 2732 |
| 10. franchise tax board : 3960 | 20. minimum wage : 2658 |

searchMaster@dts.ca.gov

New Search Engine

- Google Search Appliance
 - Model GB-1001
 - Purchased Document Limit: 2 million
 - Box limit: 3 or 5 million documents
 - 3 units, (1 hot swap), load balanced / redundant
-

We did not buy the **Google Mini**

- Hosts: 50k – 300k documents
- \$3,000 – \$9,000
- Capacity too low for DTS
- Limited configuration options



Google Search Appliance (GSA)

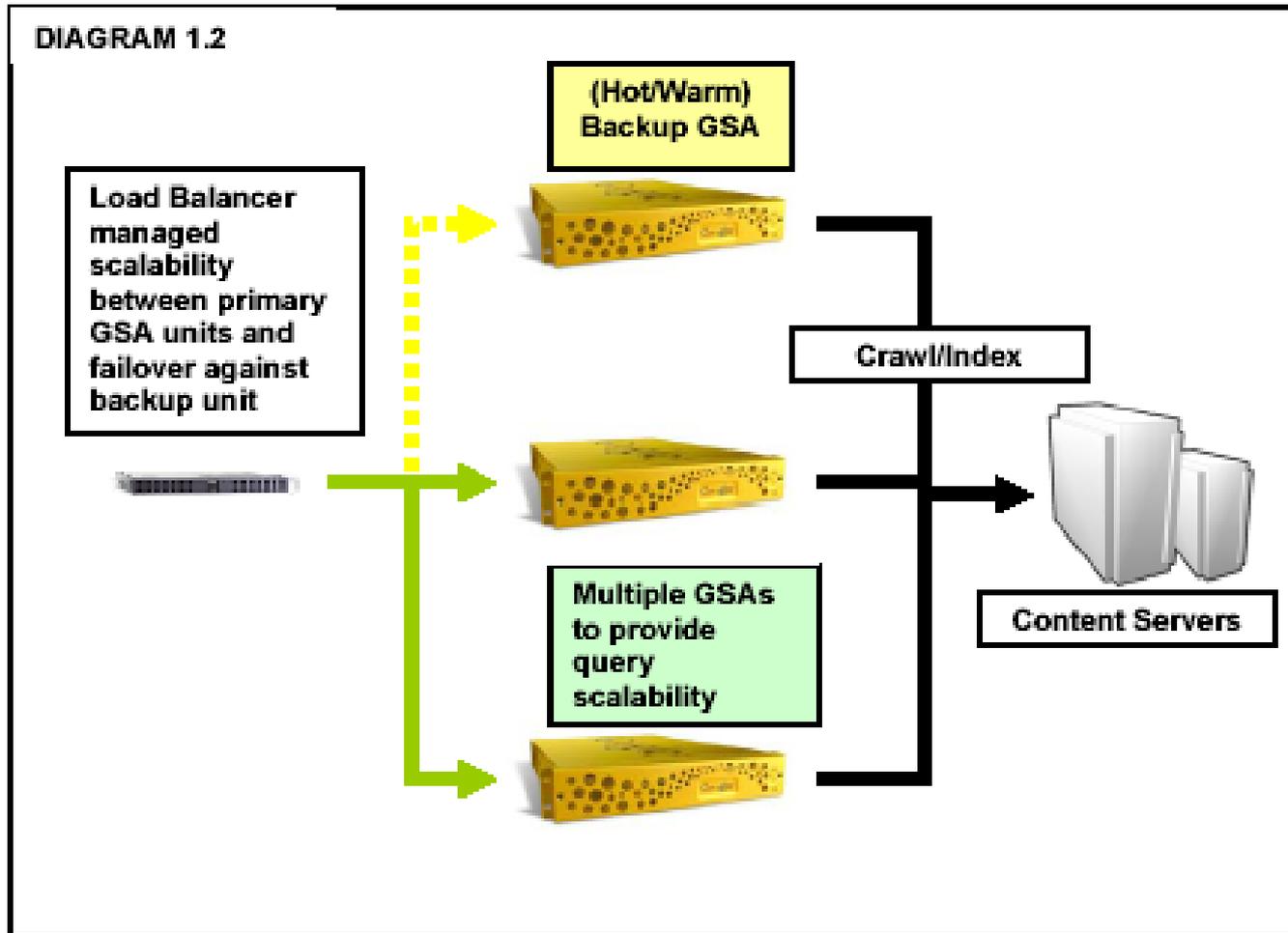


- **GB-1001 (3 each)**
- **Box Limits: 500k to 3M or 5M documents**
- **Most of the search options and features of GW:**
 - inurl, intitle, site, link, +, - , etc.
 - Spell suggest
 - Sort by
- **KeyMatch (a.k.a. Recommended Links)**

- **What's an appliance?**
- **It's all this stuff in one box:**
 - OS (some flavor of Linux)
 - Server hardware (a custom Dell PowerEdge)
 - WebServer (Apache?)
 - Application Server (Tomcat?)
 - Search Engine Software (Google Proprietary)



DTS GSA Architecture



Google Growth Path

- GB 8008
- Unlimited documents
- Lots of \$\$



GSA Implementation Status

- Took delivery August 30, 2006
 - Indexed over 2M documents
 - we're at the limit now!!
 - Content excluded due to document limit constraints:
 - a. the entire .edu domain
 - b. California county sites.
 - c. California city sites
 - d. private and non-profit sites
 - Please inform me if you discover any inappropriate content in search results
 - Is the GSA indexing your entire site? How to find out.
 - Crawling analysis to remove “garbage” URLs – on going.
 - Webmaster implementation of “Search Best Practices” – a necessary component to our success.

 - Google Portal search will launch with Portal redesign.
-

Questions So Far?

OLD

K2



NEW

GSA

Search Best Practices

for Webmasters

- What's our goal?
 - The most relevant search results possible for our visitors.
 - Did they find what they were looking for?
- Two (actually 3) Parts to Success:
 1. Superior Technology – Task Complete
 - Selection committee: PRP
 - Purchase Approved by: Executive Management
 - Install, Configure, Maintain & Tune: Search Administrator
 2. Content Preparation – follow Best Practices
 - You influence managers, content providers and other webmasters – Content Preparation Matters!
 3. Governance (outside presentation scope)
- Let's demonstrate the need for Search Best Practices in content preparation...

Governance

(sorry, I couldn't help myself)

- What's your TLD?
 - .ca.gov?
 - Are you a State of CA department?
 - Are these?:
 - <http://www.auburn.ca.gov/>
 - <http://colma.ca.gov/>
 - <http://www.hollister.ca.gov>
 - <http://www.placer.ca.gov/>
 - <http://www.sdcounty.ca.gov/>
 - <http://bigfun.org>

 - Who is the primary domain authority for the .ca.gov domain? Hint: it's not Canada.
-

Content Preparation

Demo Part I

- Google Web:
 - [search: ca dts strategic plan](#)
 - MSN Web:
 - [search: ca dts strategic plan](#)
 - Portal K2:
 - [search: dts strategic plan](#)
 - Portal K2 (Never Deployed):
 - [search: dts strategic plan](#)
 - New Portal Google Search:
 - [search: dts strategic plan](#)
-

Search California

[Next>](#)

[PDF] [Directors Letter](#)

... Establish the initial **DTS Strategic Plan** to articulate the Department's **strategic** direction and five-year road map to guide decision-making. ...

www.dts.ca.gov/files/DTS_Strategic_Plan.pdf - 2006-08-08

[Consolidation Management Office - Department of Technology ...](#)

... We are pleased to announce the recent release of the first **Strategic Plan** for the Department of Technology Services (**DTS**) and would like to thank the talented ...

www.dts.ca.gov/cmo/bulletin.htm - 22k - 2006-06-05

[Consolidation Management Office - Department of Technology ...](#)

... No. It looks completely different. We are more focused on **DTS** as a new department. The old **strategic plan** was more for a data center. ...

www.dts.ca.gov/cmo/faq.asp - 25k

[[More results from www.dts.ca.gov/cmo](#)]

[PDF] [Update to the 2004 Plan](#)

... **planning** is reflected in this updated **strategic plan** document. We also successfully created the Department of Technology Services (**DTS**) through reorganization ...

www.cio.ca.gov/PDFs/IT_Strategic_Plan_R2.pdf - 2005-11-18

[PDF] [California State Information Technology Strategic Plan](#)

... **planning** is reflected in this updated **strategic plan** document. We also successfully created the Department of Technology Services (**DTS**) through reorganization ...

www.cio.ca.gov/PDFs/101405_ITStrategicPlanUpdate.pdf - 2005-10-14

[[More results from www.cio.ca.gov/PDFs](#)]

[PDF] [New & Improved Project Management](#)

... What problems are solved? Is there a **strategic plan**? Page 26. ... What problems are solved? Is there a **strategic plan**? What is the leadership capacity to ...

www.dts.ca.gov/Training/pdf/ei/Mayer-CIONapaPresentationMarch2005.pdf - 2005-11-01

Content Preparation

Demo Part II

- “governor”

- <http://www.ca.gov/state/search/portal/search.jsp?q=governor>

- “California”

- <http://www.ca.gov/state/search/portal/search.jsp?q=California>

- “Californai”

- <http://www.ca.gov/state/search/portal/search.jsp?q=Californai>

Search Best Practices for Webmasters

- Source: 5 page document
 - written by yours truly
 - 12 Best Practices gleaned from others
 - Will be emailed with slides to all State of CA webmasters.
- Let's look at a few...



1. Use a Robots.txt File!

and the meta “ROBOTS”

- One of the best ways to spend your value time. Implement this feature.

Robots Tag: `<META name="ROBOTS" content="NOFOLLOW,NOINDEX">`

Robots.txt file format:

User-agent: *

Disallow: /calendar/

Disallow: /cgi/

- Syntax Error will disable the functionality:
“User-agnt”
-

2. Write well-structured, standard XHTML

- Google likes these three things:
 - Pages referenced by other pages
 - Well-structured, top/down design
 - Unique content in paragraph form, strong nouns, noun phrases and verbs.
- Best achieved along side your 508 and WCAG Accessibility Compliance efforts
- This is the best web site



2. Write well-structured, standard XHTML cont.

- Navigate to <http://www.csszengarden.com>
 - Disable Style Sheets
 - Voilá
 - That's the way a search engine will see your pages.
 - Tables not used for structure
 - Meaningful Content...
 - Wrapped in Perfect Structure
-

3. Validate Every Page

- Don't be careless. Validate every page XHTML STRICT – the new State standard
 - Don't forget your CSS too.
 - Lots of validation sites
 - Including: <http://www.w3.org/QA/Tools/> and <http://validator.w3.org/>
-

4. Help Content Providers

- Put a unique Title on all your pages and documents
 - Populate the Description meta-tag
 - Spell check EVERY document
 - Avoid Using Frames
 - Use a content management system
 - Make web pages for users, not search engines.
 - Forget SEO tricks with one exception.
-

5. Make your sites easy to navigate.

- Establish clear hierarchy of hypertext links
 - Offer a HTML site map
 - Consider Google's Sitemap Webmaster tool:
 - <https://www.google.com/webmasters/sitemaps/docs/en/about.html>
 - Helps speed up discovery of your pages
 - You will need a Google account
 - Has nothing to do with CA Portal search
(just thought you might like to know)
-

More?

- Yes
- No



6. Interlink (or cross-link) from other sites

- The only SEO-like effort you should consider
 - Google's Page Rank algorithm ranks pages higher if they are linked to by other sites
-

7. Delete old files

- Crawlers find pages without links
 - Possible security problem here?
 - Indexing irrelevant content makes you look bad in search results and...
 - Wastes the crawler's time and...
 - Wastes your bandwidth on crawling...
 - Uses up our precious GSA document count
 - At least use the `ROBOTS.txt` file to exclude old, archived or unneeded documents
-

8. Follow File Size Limit guidelines.

- No HTML pages over 2.5MB
 - No non-HTML (PDF, etc.) over 30MB
 - Keep PDFs under 5MB for those poor folks still on dial-up.
-

9. Get Listed

- If we aren't spidering your sites...
 - We don't know your homepage URL
 - Check the Agency Index at www.ca.gov
 - You in there at least once?
 - If not, get listed. Call or email the Portal webmaster.
-

10. Register as a California Site.

- Spidering sites would be a piece of cake if every State web site had a .ca.gov name
 - If you don't, get one. Contact the .ca.gov domain authority. Almost nobody uses state.ca.gov.
 - From [\(DTS\) CSGNet](#) website:
“Historically, CA.GOV domains *may be requested by almost any type of state government or associated entity*. However, the STATE.CA.US domain is restricted to official state government entities.”
-

Questions?



Thanks for Watching

- Contact:

- Kevin Paddock, Search Administrator
 - DTS: 464-4233
 - searchMaster@dts.ca.gov
-