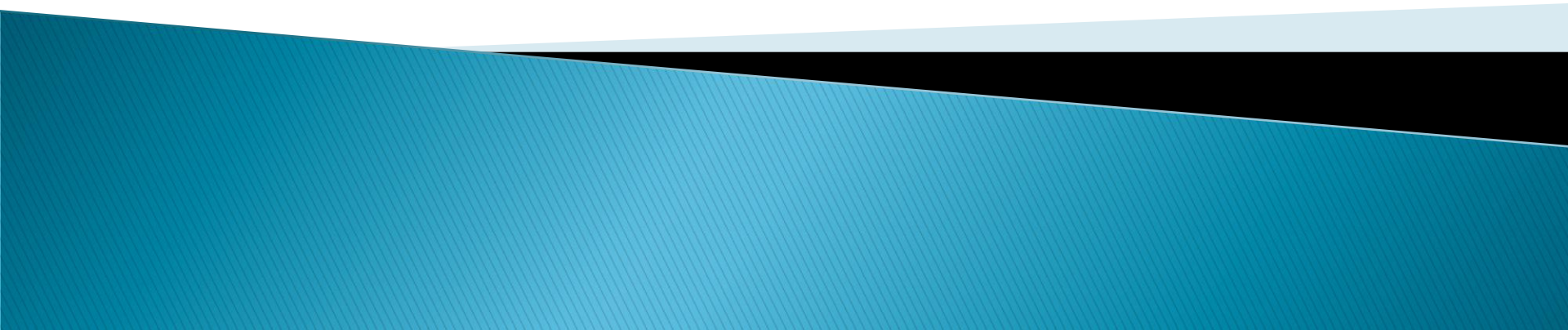


# Google Search Appliance

Deployment and Administration



# What is the GSA?

- ▶ The Google Search Appliance delivers highly relevant, fast, easy-to-use search results to users across all corporate content including:
  - File servers
  - Web servers
  - Document, CMS, Sharepoint
  - Enterprise Applications
  - <http://omni.musc.edu>



GB1001

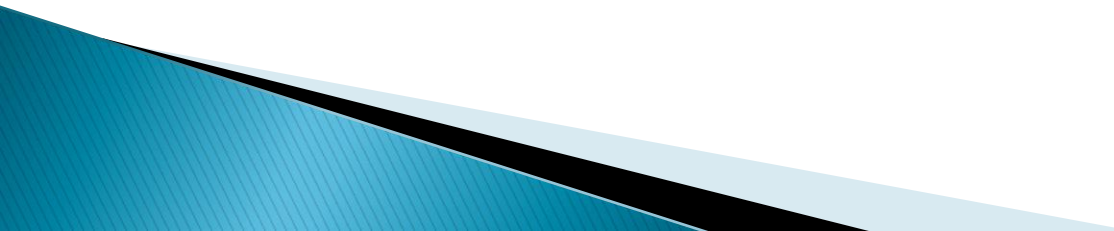


GB5005



GB8008

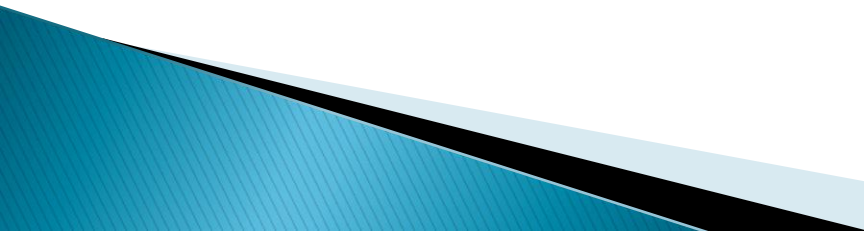
# How does MUSC use the GSA?

- ▶ Two servers behind load balancing router.
    - Primary: Google5.mdc.musc.edu (Primary)
    - Secondary: Google4.mdc.musc.edu (Testing)
  - ▶ GSA configuration set to crawl the MUSC.edu and MUSCHealth.com domains.
  - ▶ Vanity domains also crawled.
- 

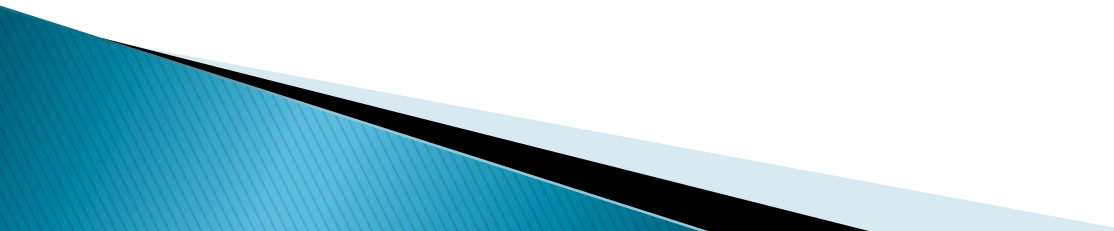
# Working with the GSA.

- ▶ Modify Google4 to test internally:
  - <http://google4.mdc.musc.edu:8000/> (Admin)
  - <http://google4.mdc.musc.edu/> (Users)
- ▶ Once testing has been accomplished, port configuration over to Google5
- ▶ Administrative port: 8000

# Collections

- ▶ Collections are subsets of the index used to serve different search results to different users.
  - ▶ Can be organized by geography, product, job function, and so on.
  - ▶ Collections can overlap, so one document can be relevant to several different collections, depending on its content.
  - ▶ Collections also allow users to search targeted content more quickly and efficiently than searching the entire index.
- 

# User Interface

- ▶ Controlled by “Front Ends”
  - ▶ Front End is XSLT style sheet.
  - ▶ Creation process for Front Ends is a one-way street.
  - ▶ Create a front end for every search type or specific searches ?
- 

# GSA UI Features

## ▶ Related Queries

- Unlike spelling suggestions, related queries are not available by default. You can create them for a specific front end by associating a search term to a related query.

## ▶ Keymatch

- KeyMatches are preferential search results, or recommended links, that appear at the top of the search results. Like related queries, KeyMatches are results that are based on the end user's original search terms.

# GSA UI Features cont.

## ▶ Narrowing Searches

- The GSA can narrow searches by providing dynamically formed subcategories ("dynamic result clusters") based on the results of each search query.

## ▶ Widening Searches

- Without any input from the end user other than a search term, the GSA can expand a query by adding synonymous terms. This helps end users get results that they would otherwise miss. The feature is called "query expansion."

# Deployment Process

- ▶ Use Google Web Master Tools to resolve outstanding issues on both .com and .edu
  - Meta-keywords & Meta-description
  - Titles of all pages
- ▶ Determine appropriate categorization for GSA appliance by querying interested departments and parties.
- ▶ Supply list of related queries, misspellings, medical acronyms, etc to improve user's search capability.
- ▶ Collaborate across MUSC verticals to provide fulfilling and functional search capabilities to all interested parties.

# Virginia.edu Example

<http://www.virginia.edu/search/>

- ▶ Features
  - Refine Results (collections)
  - Finalized header & footer
  - Appropriate categorization of all school material.
- Necessary Steps for mimicry
  - ▣ Finalized header & footers
  - ▣ Categories for GSA collections
  - ▣ Incorporate into .edu and .com domains.

# Federated Search ?

Maybe . . .

- ▶ Medtronic and the Google Search Appliance (pdf)

- ▶ Using the Google Search Appliance for Federated Searching: A Case Study

## Medtronic and the Google Search Appliance



### EXECUTIVE SUMMARY

**Challenge**  
Prior to implementing Google, Medtronic used a variety of data queries to find different types of information on its intranet, such as medical journal articles, internal presentations, and press releases. It became imperative that the company have a unified search solution across its intranet that could handle both PDF documents and queries of varying lengths.

**Solution**  
The Medtronic IT team reviewed Google's reputation for fast and accurate web search, and felt confident that the Google Search Appliance could provide the same high-quality results in its intranet.

**Product**  
Google Search Appliance (GS-1003)

**Benefit**  
"This line to return results in parentheses. I laugh when I see '()' in the browser. We went from an hour to searching to a 40 second search to get the top page."

Nathan Ahlstrom  
Principal IT Technologist,  
Medtronic



As a leader in medical technology with annual revenues of approximately \$9 billion, Medtronic enables medical professionals to develop comprehensive solutions to its patients' life-limiting disorders and chronic diseases such as heart disease, neurological disorders, and vascular illnesses. Each year, 2.5 million patients benefit from their work.

To achieve breakthroughs, Medtronic's 30,000 employees, especially research scientists and engineers, must be aware of one another's projects by having access to important internal information, such as published papers, recent medical journal abstracts, and presentations.

### The Challenge

Medtronic relied on disparate SQL search queries on its intranet, which comprise approximately 90,000 HTML, PDF, and Microsoft Office documents hosted on one server. The existing search method proved extremely unpopular and unreliable, allowing users to enter only one search term at a time and returning low-quality results. Medtronic employees were growing angry and frustrated with having to look up information in separate queries, and getting irrelevant or inaccurate results in return.

The importance of linking Medtronic employees to internal information they need cannot be overstated. The technical and research staff must have access to the latest data and findings to be able to develop cutting-edge products and treatments for their customers. It became the responsibility of Nathan Ahlstrom, Principal IT Technologist in the research division at Medtronic, to provide a unified search solution for all employees.

"A new search solution had to be able to search across our entire website and handle variable length queries. I was looking for technical knowledge management that avoided duplication of work, and encouraged collaboration across the different offices and groups within the company," recalls Ahlstrom.

"Prior to getting the Google Search Appliance, our intranet was primarily made up of a technical and medical news library, and a technical news portal. We wanted to enhance employee collaboration by developing an internal online newsletter that would focus on articles about employees and their projects, but first needed a unified way for everyone to search. The idea behind constructing the human interest articles and employee bios is to avoid the duplication of work, so people would know who and where to go to get the latest information on a topic, and be able to view employee presentations and published articles," he says.

During this knowledge management redesign, scientists and engineers had to have access to the latest medical journal and technical news articles on the company intranet. And if key information such as new disease findings or recent technological developments is missing, they can post what they know quickly and easily.

## Using the Google Search Appliance for Federated Searching: A Case Study

### Abstract

This article discusses an experiment by the University of Nevada, Reno to do federated searching of digital collections and vendor databases using version 4.1 of the Google Search Appliance (GSA). The digital content included in this test came from the university's locally held CONTENTdm and geospatial data collections and a sample of records from EBSCO's Academic Search Premier database. The latter set of records revealed many of the limitations that the GSA has in being able to successfully index and retrieve content that is dynamically generated and that requires authentication by a third party. Finally, the article briefly touches on conversations that UNR has had with the members of the New England Law Libraries Association (NELLO), which is doing a test of version 4.2 of the GSA on one of its member institutions digital collections.

### Keywords

Academic Search Premier, dynamically generated content, EBSCO Academic Search Premier, Google Search Appliance, XML

### Acknowledgements

The author wishes to thank the following individuals for their assistance with the project discussed in this article:

Steven D. Zink, University of Nevada Reno  
Araby Greene, University of Nevada Reno  
Gregg Stefanelli, University of Nevada Reno  
Rick Anderson, University of Nevada Reno  
Tracy L. Thompson, NELLO  
Roberta Woods, Franklin Pierce Law Center Library  
Tamara Scott, Google Enterprise Search  
John Gregory, Google Enterprise Search  
Michael Gorrell, EBSCO Publishing

### Author

Mary Taylor ([tytormk@unr.edu](mailto:tytormk@unr.edu)) is the Metadata Services Coordinator for the University of Nevada, Reno.

### Introduction

As the Metadata Services Coordinator for the University of Nevada, Reno, two of my areas of interest are metadata harvesting and federated search. During my first months in this position I served on a committee to review and evaluate federated search products. After numerous meetings with vendors and product trials, the committee came away with the conclusion that the functionality offered by these products did not meet the price quotes we were receiving. When I first started in this position and inquired about the five most important tasks that I should be working on, the Dean of University Libraries, Dr. Steven D. Zink, mentioned that finding out the costs and technical requirements for implementing the Google Search Appliance (GSA) should be at the top of the list. The GSA is

# Next Steps?

Get house in order. . .

- ▶ Categorize MUSC's own resources in the GSA first.
  - Web (Site Exec, SharePoint, Policies, etc)
  - Databases
  
- ▶ Game plan for Federated Search
  - Statement of Work
  - Goals & Milestones
  - Implementation Plans
  - Software Engineering