

# Lunch and Learn

Seven Strategies for “Search”

A.J. Million  
January, 19<sup>st</sup> 2010  
Organizational Results



This presentation does not provide rules. Only strategies...

# The “Deep Web”

## What is it?

The Internet is not as “open” as we think. Only some files are freely available. Others are just accessible in specialized databases.

Google, Bing, and most search engines can only index open web directories. That’s [about 20%](#). The rest is considered the “Deep Web” and can include:

- Material from specialized databases.
- Data without keywords for identification.
- Files on local area networks.
- Protected files on web servers.

# The “Deep Web”

## What this means?

When you use a search engine only publicly accessible files will come up. [Search engines](#) can't communicate with specialized databases or index private file directories.

A good searching strategy is to start using a search engine since it's quick and easy. Next, if the right results don't come up, then check everywhere else.

Images are hard to find because they don't have characters that search engines can “match” against.

Knowing what is public is a major consideration of search.

# RSS Feeds

## What are they?

RSS stands for “Really Simple Syndication” feeds. They are links to web addresses and let you set up specialized data streams based on what you want.

They are identifiable by an orange logo with lines coming from a center dot. They also update in real-time. You can use them to avoid checking for multiple updates.

Feeds are composed of a special type of code (XML) that any computer can read.

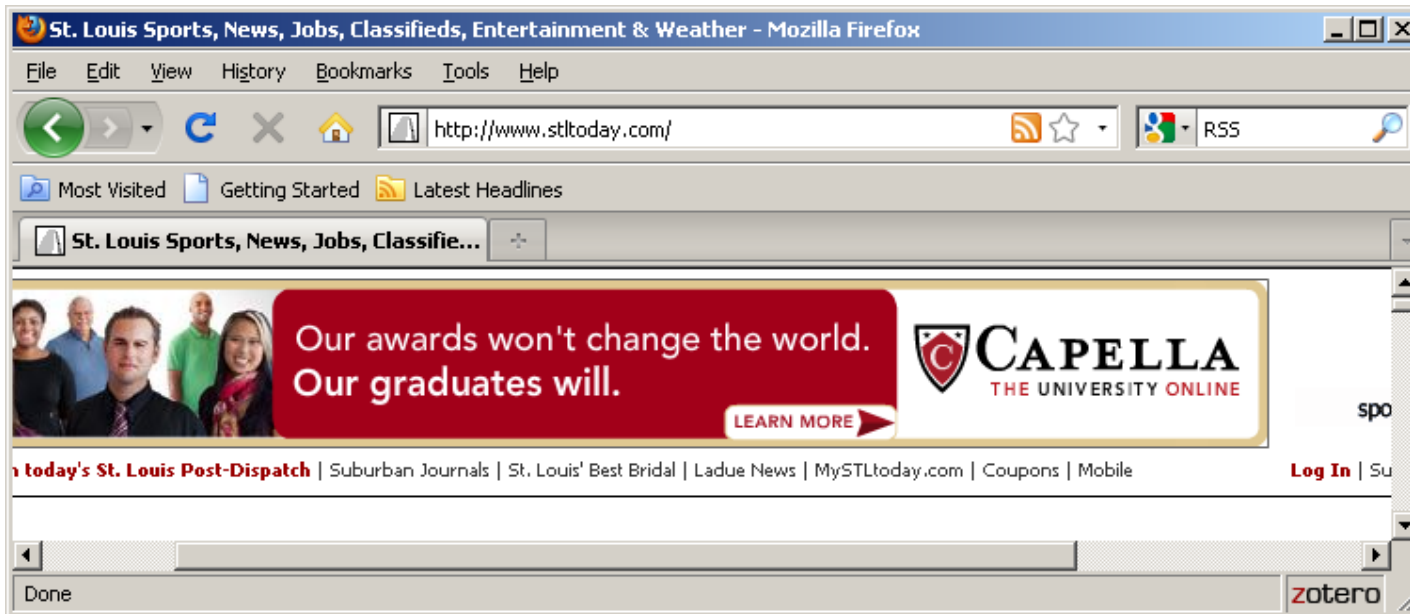
Windows 7 can stream them to your desktop.



# RSS Feeds

## How can you use them?

There are several ways to use RSS feeds, but popular ones include feed readers, web browsers, and operating systems.



# Special Search Operators

## What are they?

Most search engines operate using a complex set of algorithms.

They match keyword characters to identify what users want to find.

In special situations, specific [commands](#) make more sense to use than “plain language.” These shortcuts let you manipulate search results.

Operators vary, but there are a few common ones that are especially useful. Their value is mostly selective and not worth using all the time.

# Special Search Operators

## How can you use them?

Phrase Searching: By putting quotes around a phrase [i.e. **“bituminous pavement in Missouri”**] you tell the computer to find that exact set of words. This is good for titles.

Site Specific: If you search with a keyword and site query [i.e. **asphalt site:modot.mo.gov**] you eliminate others. Taking this approach is great if you only want results from one website.

Exclude Word: To get rid of an improper keyword add a minus before it [i.e. **traffic merging -late**] to eliminate returns. A good time to use this is when a phrase has dual meanings.

# Social Media Software

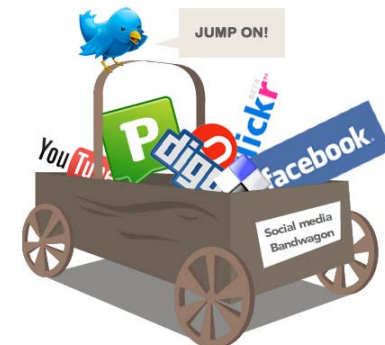
## What is it?

Twitter, Facebook, MySpace are online tools that enhance communication. They aren't normally considered to be "search" tools but still offer an alternative to query based systems.

How? Think of social media as a way to digitally ask someone in your network a question. They're great for following trends.

Some tools that can be useful include:

- Facebook, Linked In, YouTube
- Twitter, Flickr, Digg, del.icio.us



# Social Media Software

## What can some of them do?

Twitter: <http://www.twitter.com>

Lets you set up feeds, send them to others, etc. As a research tool it emphasizes “finding” over “searching.”

Facebook: <http://www.facebook.com>

A very general site, Facebook lets you do a little of everything. It's good for staying in touch and communicating with others.

Digg: <http://www.digg.com>

Basically, Digg is your website bookmark list on steroids. It provides multiple ways to manage links to what you've already found.

# Cloud Computing

## What is it?

How many of you remember the old computer terminals that you used to use? Cloud computing is similar.

Because storage is cheap and the internet is so fast now there's a movement to run basic applications and store data online. The internet is moving that direction.

Nobody is sure where Cloud Computing will lead, but online publishing tools help hint at what's to come.

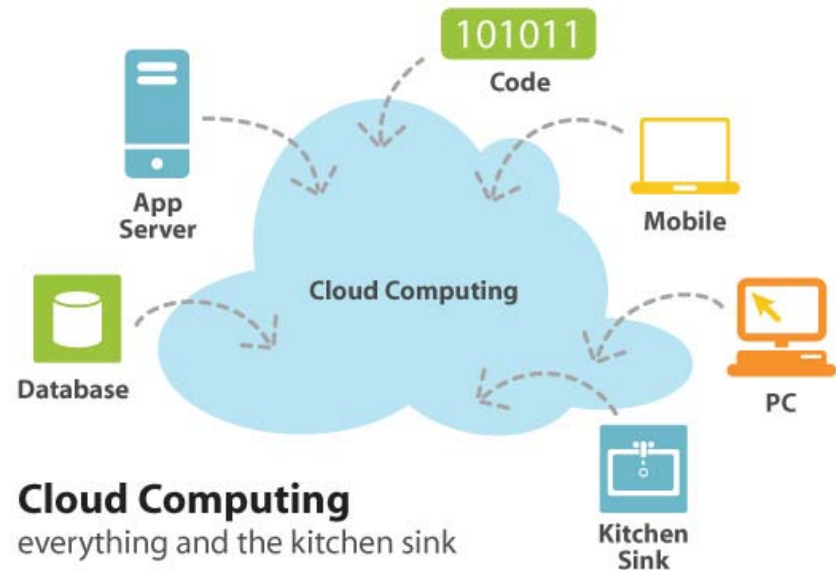
A 2007 [IBM whitepaper](#) predicted that data will double online by 2011!

# Cloud Computing

## How does this relate to search again?

As more and more “stuff” moves to the internet, much of it is going to become part of the “deep web.”

You can exploit this! If you identify places you routinely visit you'll have more success than if you search *everywhere*. Don't cast a wide net unless you don't know what you are looking for...



# Controlled Vocabularies

## What are they?

Controlled vocabularies are pre-defined subject headings used in the place of natural language. They help eliminate redundancies and double-meanings.

Terms vary, but most databases offer special subject headings to users. They cut out problems inherent in language.

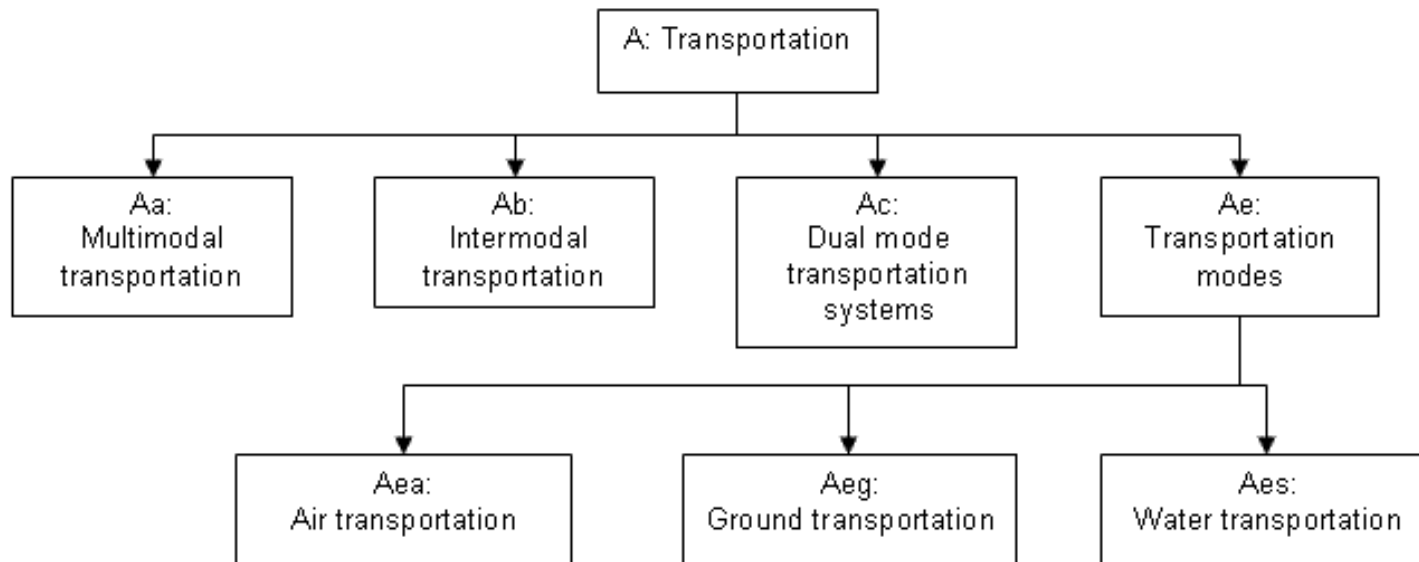
A few examples include :

- “Cats” and “Felines.” Substitute “Cats” instead in subject.
- “Maine Coon.” Along with “Calico” is a defined as a “cat” type.

# Controlled Vocabularies

## How are they used?

Hundreds of controlled vocabularies exist but most are subject specific. The [Transportation Research Thesaurus](#) is used by TRB and a relevant example.



# Pathfinding and Usability

## What is it?

Pathfinding and site usability refers to design. Well designed websites and databases all *share features* that make them easy to navigate.

These features operate in the same way that road signs do. If you know what a certain feature means – or does – then you'll find your way much easier.

Since this is such a broad topic there's no way to know everything. I'll limit my discussion to TRIS.

Many design elements are universal.

# Pathfinding and Usability

## What do I look for?

Breadcrumbs, e-mail features, controlled vocabularies, and check boxes... All of these are used for a reason.




TRIS > **Search Results**

0 Marked Records: [Print](#) | [Email](#) | [Save](#) | [View](#) | [Clear](#)

You searched on "missouri" in keywords


Limit Results   [Modify](#) | [New Search](#) | [Search History](#)

Showing 1 - 25 of 1254 records  of 51 >>

 **RSS Feed**    

[Mark: All](#) | [None](#) | [Show](#) [Record Summary](#) [Published](#)

---

**State of Bridges. Shockingly High Number of Bridges Remain Sub-Standard** 2009/11/1  
*Better Roads*, 2009, pp 8-19  
Actions: [View Record](#) | [View Full Document](#) 

---

Showing 1 - 25 of 1254 records  of 51 >>

[Home](#) | [Filter search](#) | [Modify search](#) | [Help](#)

# Quick Takeaways

## **What Should I remember?**

1. The “deep web” means Google can’t find everything.
2. RSS feeds can be a useful way to bring information to you.
3. Special commands offer extra flexibility.
4. Social Media is a quick way to way to “ask a friend.”
5. Controlled vocabularies standardize language.
6. Cloud computing means check portals.
7. Design is king! Learn the way things are formatted.