

SEO

<dit>



DITA + SEO:

How and Why You Should Optimize DITA-based Content for Search Engine Optimization

Keith Schengili-Roberts – July 30, 2019



Scope of this Presentation

- Focus: a *practical* approach to SEO for DITA-using technical writers publishing HTML content to the web
- Along the way I *may* burst a few bubbles when it comes to what techniques do and do not matter
- What I will be looking at:
 - SEO and how search engines work (briefly)
 - HTML-based output from DITA content
 - Mechanisms available in DITA to aid with SEO
 - Information on what Google is looking for when it ranks content
 - Writing DITA content with better SEO in mind
- And along the way talking about how Zoomin optimizes some of these processes



Who's This Guy?

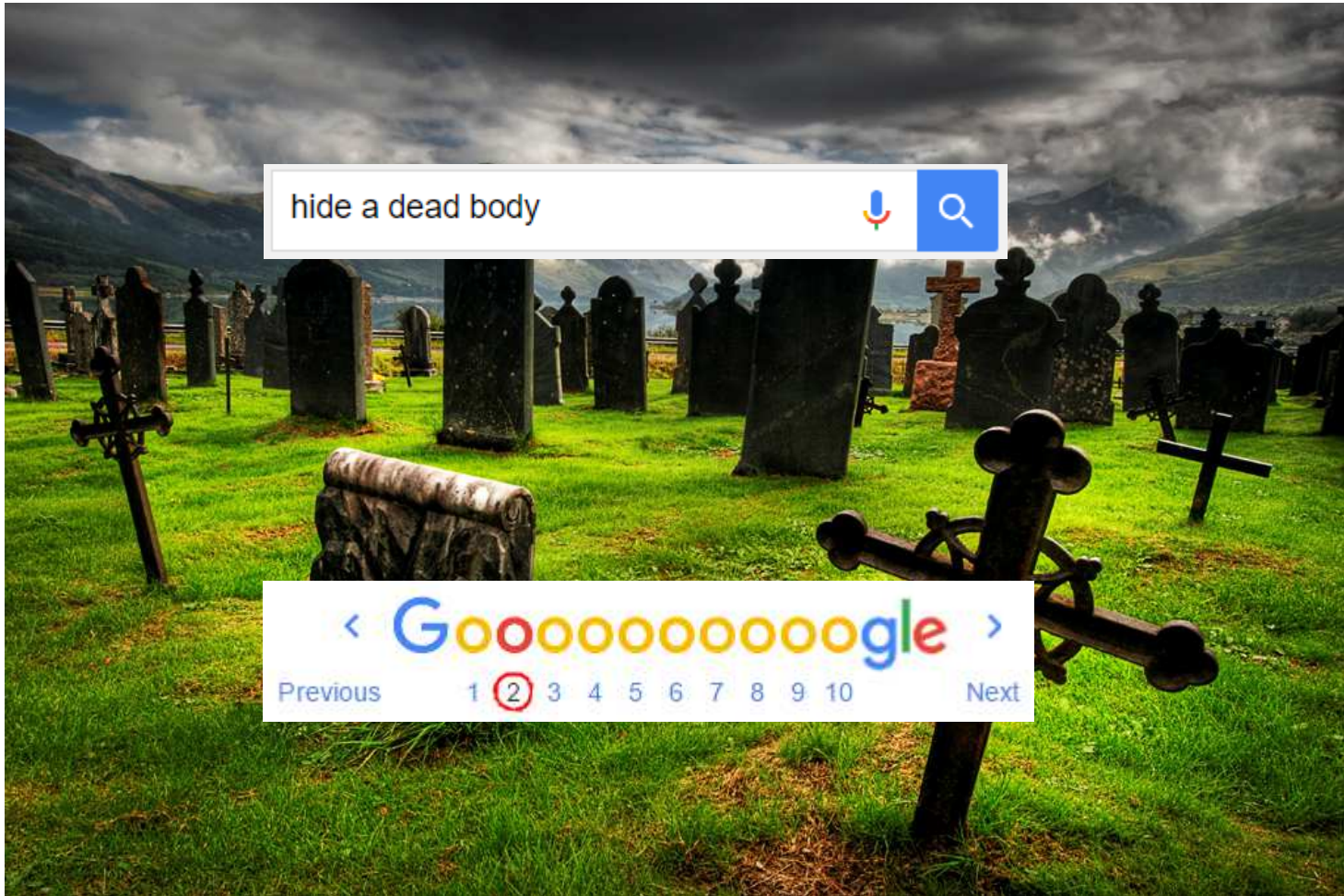
Keith Schengili-Roberts

- Working in Tech Comm since early 1990s
- Working with DITA since 2004 (pre-DITA 1.0)
- Senior DITA Content Strategist at IXIASOFT
- Chair of OASIS DITA Adoption Committee, member of LwDITA SC and DITA TC
- The person behind DITAWriter.com





Q: Where's the best place to hide a dead body?

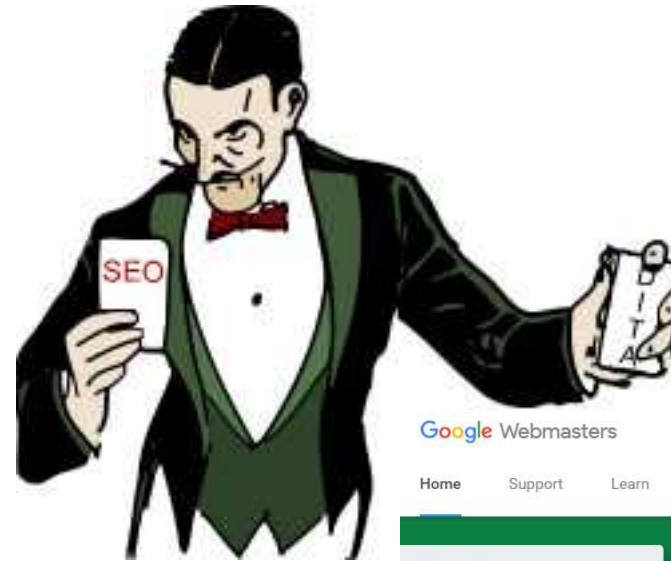




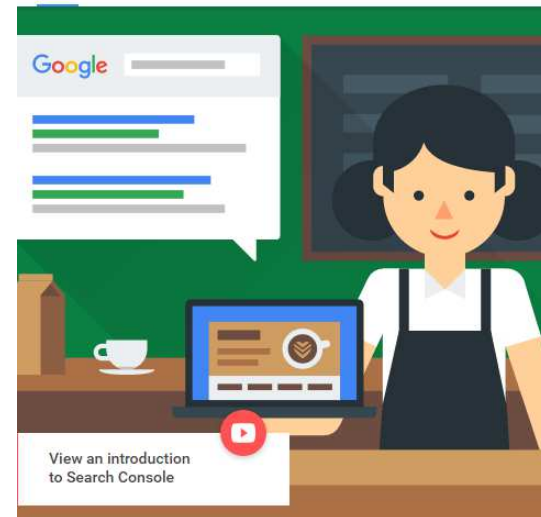
Search Engine Optimization is Magic!

- No, it really isn't
- There are agencies that can help with SEO, but the information *is out there and available*
- Recommend Google Webmasters as a start:

www.google.com/webmasters/



Google Webmasters
Home Support Learn Connect





Tech Docs are Not Blogs

- A lot of SEO material out there is focused on ranking marketing material and blog content; there are only a handful of articles on tech docs and SEO
- Many typical SEO recommendations may not apply or lie outside of your control
- What *is* clear is that good a corporate SEO strategy ought to factor in any technical documentation as part of an overall strategy

Domain Factors



1. **Domain Age:** In [this video](#), Matt Cutts states that:

“The difference between a domain that’s six months old versus one year old is really not that big at all.”

In other words, they *do* use domain age...but it’s not very important.

2. **Keyword Appears in Top Level Domain:** Doesn’t give the boost that it used to, but having your keyword in the domain still acts as a relevancy signal. After all, they still bold keywords that appear in a domain name.

3. **Keyword As First Word in Domain:** A domain that starts with their target keyword has an edge over sites that either don’t have the keyword in their domain or have the keyword in the middle or end of their domain.

4. **Domain registration length:** A [Google patent](#) states:

“Valuable (legitimate) domains are often paid for several years in advance, while doorway (illegitimate) domains rarely are used for more than a year. Therefore, the date when a domain expires in the future can be used as a factor in predicting the legitimacy of a domain”.

BackLinko-Google’s 200 Ranking Factors:
backlinko.com/google-ranking-factors



What Search Engines Do and Why

- Search engines want to provide users with the most relevant result to a given search query
- This is why product documentation is important from an SEO perspective, as users may find relevant answers to their queries from tech docs as opposed to a *poorly-focused* marketing communication that does not focus squarely on users' needs

High-quality technical information is a decisive factor for purchase decision

IBM has been developing a technical communications body of knowledge with a focus on content as a business asset. The company runs a permanent survey on its website asking customers to evaluate the importance of high-quality technical information. The responses of 179 participants of this survey between 2010 and 2013 have been summarized in an article for the Technical Communication Magazine of the STC.

Respondents overwhelmingly agreed that high quality technical information is important or very important with regard to the

- the perception of the company (84.5%)
- the satisfaction with the product or solution (93.3%)
- the perception of the product quality (96.4%) and
- the initial purchase decision (88.7%).

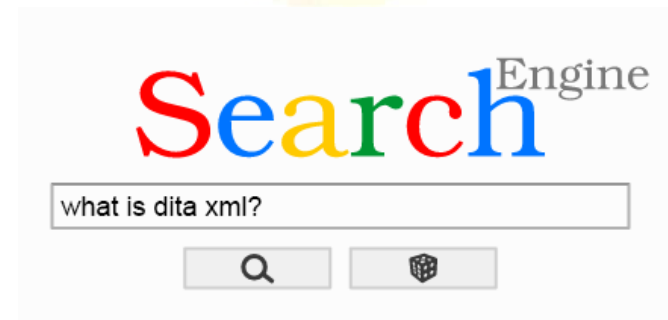
From "A business case for technical communication – facts & figures"
www.tcworld.info/e-magazine/technical-communication/article/a-business-case-for-technical-communication-facts-figures/



How Search Engines Work

Three key phases:

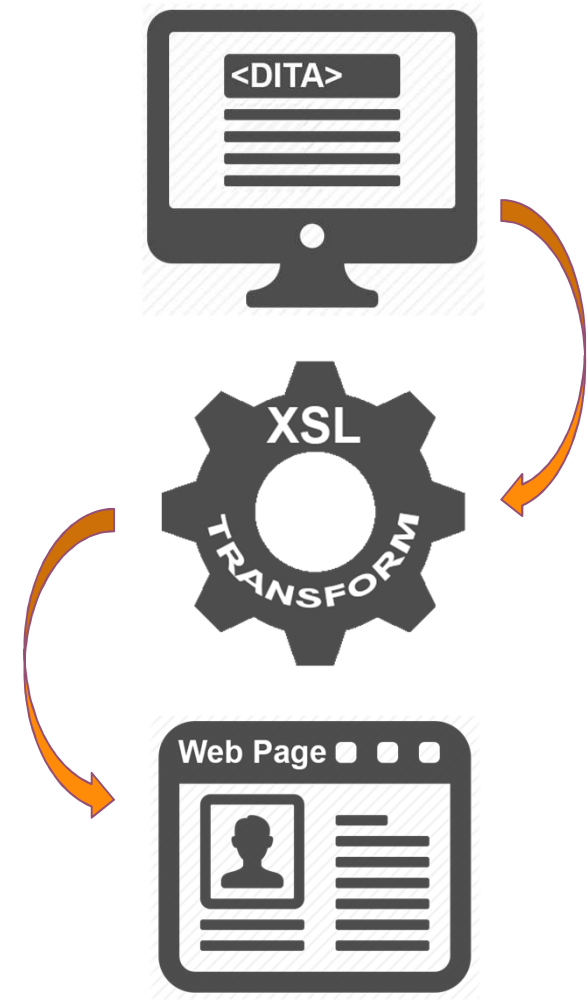
1. Your content is “crawled” by a search engine spider; finds new/changed info and retrieves it
2. Search engine analyses and indexes your website’s content
3. A user submits a query to a search engine, providing a list of possible links





How DITA Content is Produced for the Web

1. DITA content is crafted by writers
2. Content is transformed from DITA via XSL (typically through the DITA Open Toolkit) to XHTML
3. This transformed content is then placed on the Web
 - There are steps at each of these stages that can help improve SEO



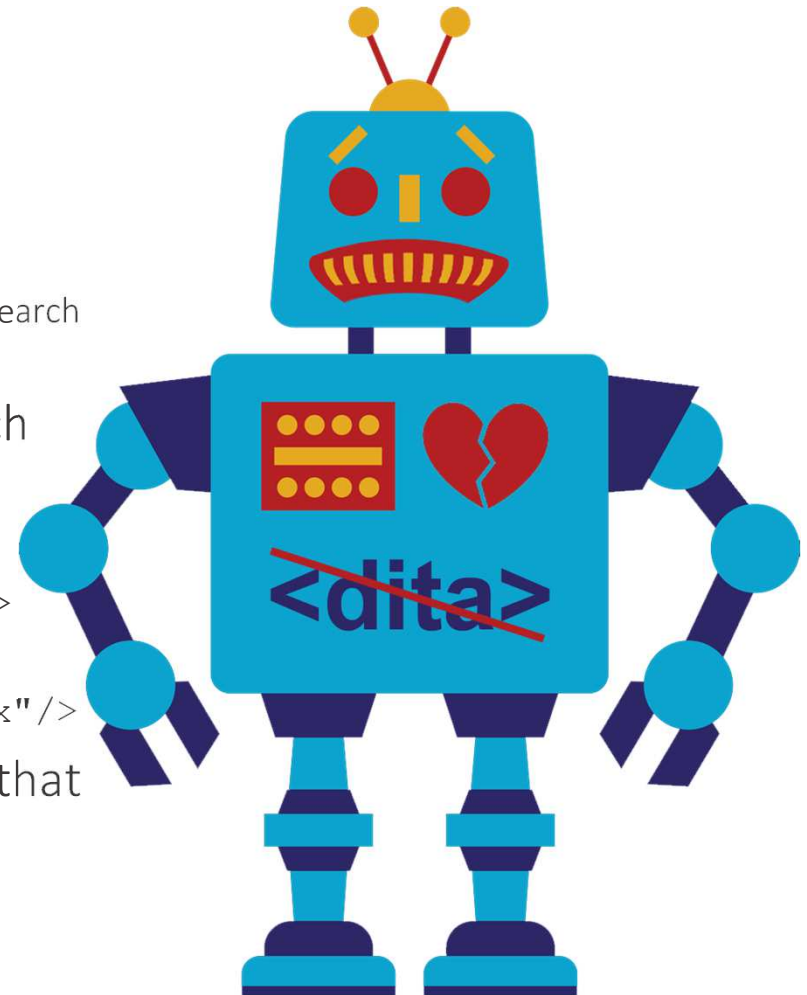


Do You Want to Even Make Your Docs Visible?

- Some companies opt not to have their documentation “spider-able”:
 - Company wants search engines to focus exclusively on marketing content
 - When there’s a need to point to a company-sponsored search engine specifically for docs
- In your webpages, add the following to each header:
 - All search engines:

```
<meta name="robots" content="noindex"/>
```
 - Google only:

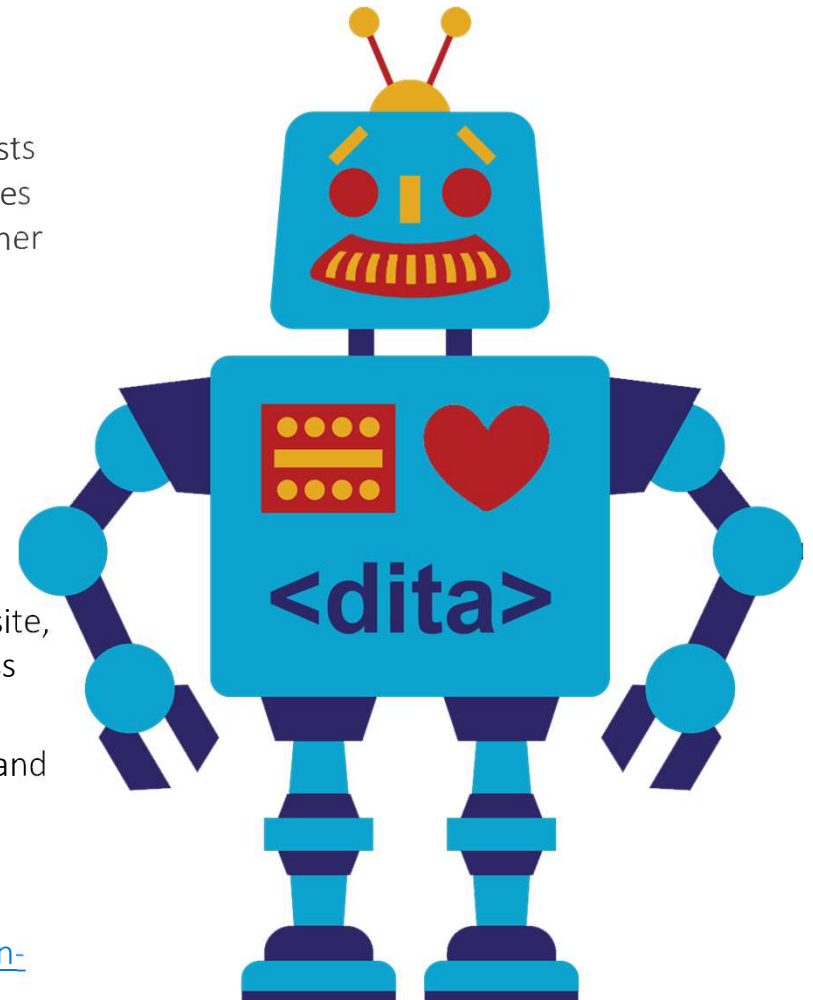
```
<meta name="googlebot" content="noindex"/>
```
- Or, add a robots.txt file to your webserver that says the following:
 - User-agent: *
 - Disallow: /tech-docs/
 - And btw, you can also set this up within Zoomin





How About Making Your Docs *More* Visible?

- Zoomin's got something for that: sitemaps
 - They are XML files that use an industry standard which lists sets of URLs for web pages within your site, how site pages are organized, and each page's priority relative to the other pages.
 - Search engines crawl these files, which tells them which webpages *you* think are important
- Zoomin provides two flavors: static and dynamic sitemap generation
 - **Static sitemap:** Manually create, maintain, and upload a single sitemap. When there are content changes on the site, these changes will not be reflected in your sitemap unless you update and re-upload it.
 - **Dynamic sitemap:** Can create one or more sitemap files, and automatically generates a new version when content is uploaded or deleted, or when a full website re-index is triggered
- For more info: https://docs.zoominsoftware.io/bundle/zoomin-configuration-guide/page/publish_ss/c_sitemap.html





DITA and Metadata

- Metadata can be incorporated at both the map and topic levels
 - Bookmaps use the bookmeta and topicmeta elements as containers
 - Topics incorporate metadata within the prolog
- This content along is then expressed at output primarily as Dublin Core metadata

```
<bookmap>  
  <booktitle>  
    <booklibrary>TRS-80 Manuals: </booklibrary>  
    <mainbooktitle>Radio Shack TRS-80 Expansion Interface</mainbooktitle>  
  </booktitle>  
  <bookmeta> ←  
    <searchtitle>A DITA-fied Example of TRS-80 Expansion Interface</searchtitle>  
    <shortdesc>This manual for a long-obsolete computer peripheral of how to port legacy content to DITA. </shortdesc>  
    <authorinformation>  
      <personinfo>  
        <namedetails>  
          <personname>  
            <firstname>John</firstname>  
            <middlename>C.</middlename>  
            <lastname>Anonymous</lastname>  
            <generationidentifier>Sr.</generationidentifier>  
            <otherinfo>Notably hard to find</otherinfo>  
          </personname>  
        </namedetails>  
      </personinfo>  
    </authorinformation>  
  </bookmeta>  
  ...
```

Bookmap

```
<title>Overview of the <term conkeyref="product_variables/computer_name">Interface</title>  
<shortdesc>What the <term conkeyref="product_variables/computer_name"/> the peripherals that can connect to it. Includes basic information on how it can be used with the <term conkeyref="product_variables/computer_name"/>  
<prolog> ←  
  <author type="creator">John Anonymous</author>  
  <author type="contributor">Jane Anonymous</author>  
  <source href="http://www.gutenberg.org/ebooks/27469"/>  
  <publisher>ACME Publishing</publisher>  
  ...
```

Topic



What is Dublin Core?

- Dublin Core is a set of metadata designed to describe web content, related to semantic web initiative
- Originating in 1995, since mid 2000s DCMI have worked with W3C on Semantic Web efforts
- DITA-OT uses a subset of Simple Dublin Core v1.1 when outputting to XHTML



- | | |
|-------------|------------|
| Title | Format |
| Creator | Identifier |
| Subject | Source |
| Description | Language |
| Publisher | Relation |
| Contributor | Coverage |
| Date | Rights |
| Type | |

**Dublin Core Metadata
Element Set Version 1.1**



DITA, DITA-OT and Dublin Core

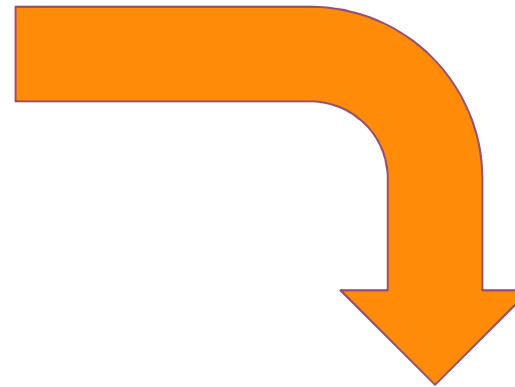
DITA Element(s)	Dublin Core Equivalent
author (topic), authorinformation (map)	Creator
category	Coverage
[output type: XHTML]	Format
critdates	Date
[id value associated with topic type]	Identifier
publisher (topic), publisherinformation (map)	Publisher
copyright	Rights
source	Source
keyword	Subject
title	Title
[topic type]	Type



DITA to Dublin Core

```
<concept id="overview_of_the_trs80_expansion_interface">
  <title>Overview of the
    <term conkeyref="product_variables/computer_name"/>
    Expansion Interface</title>
  ...
  <prolog>
    <author type="creator">John Anonymous</author>
    <author type="contributor">Jane Anonymous</author>
    <source href="http://www.gutenberg.org/ebooks/27469"/>
    <publisher>ACME Publishing</publisher>
    <copyright>
      <copyyear year="2016"/>
      <copyrholder>Project Gutenberg</copyrholder>
    </copyright>
    <critdates>
      <created date="2008-12-09" golive="2008-12-09"/>
      <revised modified="2016-05-12" golive="2016-05-12"/>
    </critdates>
    <permissions view="all"/>
    <metadata>
      <category>Obsolete product manual</category>
    </metadata>
  </prolog>
```

DITA topic



Equivalent XHTML output

```
<meta name="DC.Type" content="concept"/>
<meta name="DC.Title" content="Overview of the TRS-80 Expansion Interface"/>
<meta name="DC.Source" content="http://www.gutenberg.org/ebooks/27469"/>
<meta name="DC.Coverage" content="Obsolete product manual"/>
<meta name="DC.Creator" content="John Anonymous"/>
<meta name="DC.Contributor" content="Jane Anonymous"/>
<meta name="DC.Publisher" content="ACME Publishing"/>
<meta name="DC.Rights.Owner" content="Project Gutenberg 2016" type="primary"/>
<meta name="DC.Rights.Usage" content="all"/>
<meta name="DC.Date.Created" content="2008-12-09"/>
<meta name="DC.Date.Modified" content="2016-05-12"/>
<meta name="DC.Date.Issued" content="2016-05-12"/>
<meta name="DC.Date.Available" content="2016-05-12"/>
<meta name="DC.Format" content="XHTML"/>
<meta name="DC.Identifier" content="overview_of_the_trs80_expansion_interface"/>
```



Google and Dublin Core

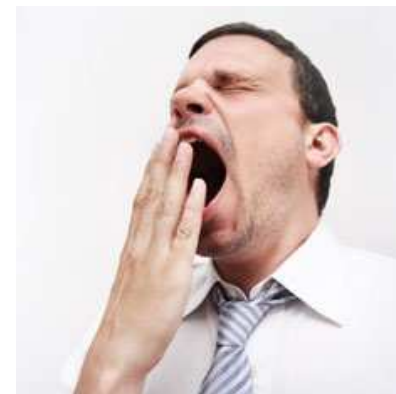
- While Dublin Core is long-established, and the DITA-OT supports it, Google does not appear to do much with this content
- It *can* be advantageous from a content management perspective
 - For example, info on when a topic is created and by whom may be useful to know
 - Local webserver may be able to filter content on DC values

Google

+



=

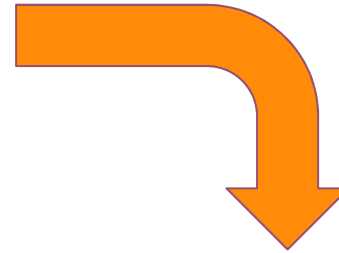




What About Metadata Keywords and Google?

```
<metadata>  
<keywords>  
  <keyword>TRS-80</keyword>  
  <keyword>Expansion Interface</keyword>  
  <keyword>Peripheral</keyword>  
</keywords>  
</metadata>
```

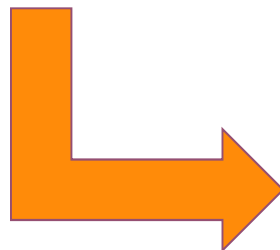
DITA topic



Equivalent XHTML output

```
<meta name="DC.subject" content="TRS-80, Expansion Interface, Peripheral"/>  
<meta name="keywords" content="TRS-80, Expansion Interface, Peripheral"/>
```

- Forget it, no point (at least from an SEO perspective)



Google Webmaster Central Blog

Official news on crawling and indexing sites for the Google index

Google does not use the keywords meta tag in web ranking

Monday, September 21, 2009

Recently we received some questions about how Google uses (or more accurately, doesn't use) the "keywords" meta tag in ranking web search results. Suppose you



So What is important to Google?

```
<link rel="canonical" href="latest-and-greatest.html" />
```

relevant_filenames.html

<title> and <searchtitle>
Short Descriptions

```
<title>IXIASOFT DITA CMS - Enterprise-scale component content management (CCMS)</title>
<shortdesc>DITA CMS is the leading component content management solution (CCMS) worldwide.
Specifically designed to manage the entire DITA documentation process, DITA CMS provides all the
tools required to "go big" and supports a complex documentation workflow from authoring to
reviewing, localizing and publishing.</shortdesc>
```



- Related concepts
- [Advanced Air Work](#)
- [Prohibited Maneuvers in the B-25](#)
- [Spins](#)
- [Dives](#)
- [Instrument Flying](#)
- [Formation](#)
- [Strange Field Landings](#)
- [Crosswind Landing](#)
- [Stalls](#)
- [Slow Flying](#)
- [Short-Field Takeoff](#)
- [Short-Field Landings](#)
- [Single Engine Operation](#)
- [Single Engine Practice](#)
- [Effect of Power and Airspeed on Single Engine Operation](#)
- [Single Engine Landing](#)
- [Tips on Single Engine Operation](#)
- [Auxiliary Hydraulic Pump and Emergency Hydraulic Selector Valve](#)
- [Tactical Uses of the B-25](#)

relationship table links
(internal links)



DITA and Filenames

- Having a relevant keyword in the webpage's filename is considered a significant signal to search engines
- In your DITA output, having a naming convention that incorporates your product name and is descriptive ought to help
- So if you make a product called "Vebulon 5" with a specific feature people might want to know more about:
 - `Introduction_to_the_product.html` ❌
 - `Understanding_the_Fliputron_feature_on_the_Vebulon_5.html` ✅



Making <title> Count

- Avoid boilerplate titles (i.e. “Introduction”); make them descriptive (i.e. “What You Need to Know About the Vebulon 5”)
- Make them concise; Google truncates long titles that are just over 70-75 characters long
- Use relevant keywords (like product name) but don’t go overboard! (i.e. “All About the Vebulon 5 – Vebulon Five, Fifth Vebulon, vebulon five, 5th vebulon, vebulon the fifth, Acme Corporation’s Vebulon Five”)



Consider using <searchtitle> as well

- In addition to using <title>, there's also <searchtitle> which can be used when you want to display alternate <title> content within a search engine

- In code, it looks like this:

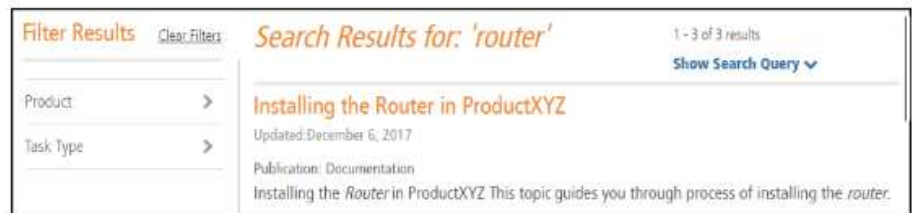
```
<title>Installing the Router</title>  
<titlealts>  
  <searchtitle>Installing the Router in ProductXYZ</searchtitle>  
</titlealts>
```

- Within Zoomin, you can see the differences in the results: local use vs. search engine

Figure 1. The Topic Page



Figure 2. The Search Results Page



https://docs.zoominsoftware.io/bundle/zoomin-configuration-guide/page/content_manage/t_search_title.html



Use Canonicalization to Point to Latest-and-Greatest

- The `<link>` tag in HTML can be used to point to a specific version of a webpage which is considered “canonical”; in other words, telling the search engines which version of a topic should appear in the results
 - Search engines tend to weigh older content more highly than new content, so older versions of techdoc topics often appear in search results
 - Using `<link rel="canonical" href="latest-and-greatest.html" />` within older versions of docs to point to most recent, which then gets used by search engines
 - This can be done automatically within Zoomin:
https://docs.zoominsoftware.io/bundle/zoomin-configuration-guide/page/content_manage/c_canonical.html

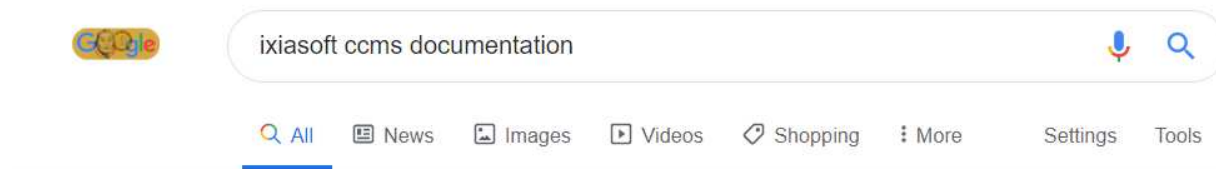


Short Descriptions and SEO

- Short Descriptions can help with SEO, as they are displayed immediately after the webpage title:

```
<title>Introduction</title>
<shortdesc><keyword keyref="CCMS_long"/> is a documentation solution that
  combines the mechanics of an extensible environment, a powerful Content Management System
  (CMS), and the elegant modularity of DITA.</shortdesc>
```

```
<title>Introduction</title>
<p class="shortdesc">
<span class="keyword">IXIASOFT CCMS</span> is a documentation solution that
  combines the mechanics of an extensible environment, a powerful Content
  Management System (CMS), and the elegant modularity of DITA.</p>
```



About 10,600 results (0.49 seconds)

CCMS Desktop User Guide - IXIASOFT

https://www.ixiasoft.com/documentation/IXIASOFT_CCMS/.../per1389985945656_2....

IXIASOFT CCMS is a **documentation** solution that combines the mechanics of an extensible environment, a powerful Content Management System (CMS), and ...

DITA



XHTML



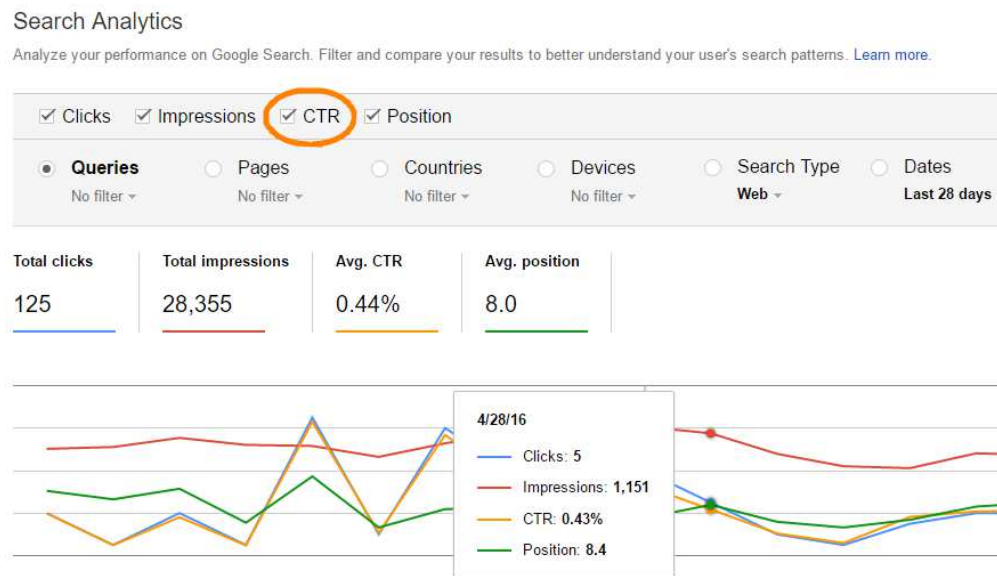
Google Search





Short Descriptions and Click-throughs

- While short descriptions are not factored in search engine *rankings*, user behaviours *are*
- Google measures click-through rates (CTR)
- A well-written, descriptive short description ensures more click-throughs





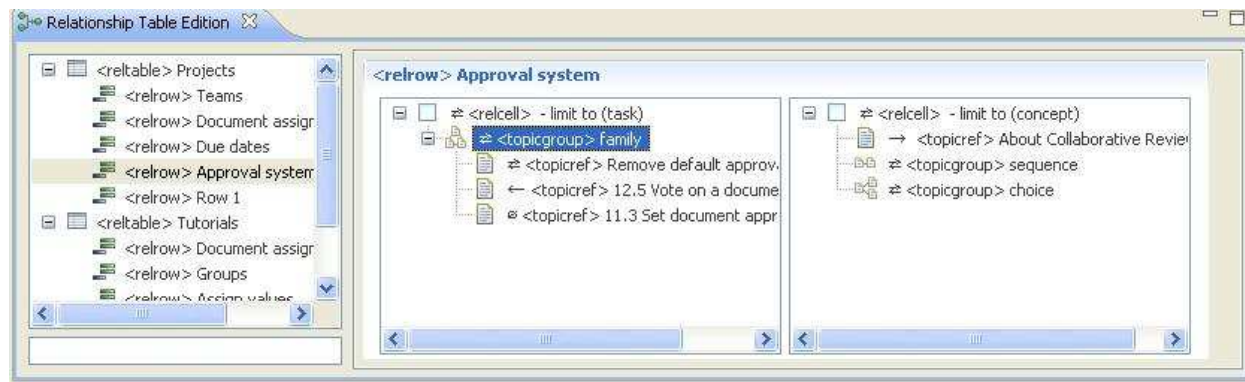
Writing Effective Short Descriptions for SEO

- A well-written short description tells the would-be reader why it is worth clicking on
 - **Task:** tell users what they can accomplish
 - **Concept:** tell users about what you are describing and why they should care
 - **Reference:** tell users what the referenced item does or what it can be used for
 - **Troubleshooting:** describe the symptoms of a problem a user may encounter and let them know that this topic can help
- While shortdesc best practices suggests two sentences, Google truncates search results at ~156 characters
 - Need to put most important content first!



Internal Links and Relationship Tables

- One metric known to influence webpage rankings are the number of internal domain links within a page
- Best of way of doing this with DITA content is to use relationship tables to build links between related topics
- So adding relationship tables is not only a DITA best practice, but it enhances SEO too!





~~Creating Killer SEO Content~~ Writing for Your Users

- Many SEO-oriented websites talk about how to “trick” the likes of Google (which, let’s face it, is probably the stupidest thing you can do), but from a technical writing perspective I would argue that it is really all about knowing your audience, being honest with them, and ensuring that you deliver the content that they need when and where they need it





A Story...



Kenmore Model 80
Clothes Washer



One day it stopped in the middle of a wash, and wouldn't drain...



So What to Do?

YouTube CA Search



0:00 / 3:06

Kenmore 80 Series won't spin or drain due to bad lid switch

rjskoko [Subscribe](#) 80

211,398 views

+ Add to Share More 382 21



211,398 views

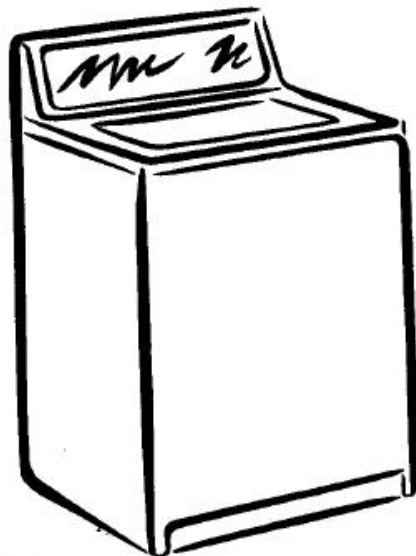


So What About the Clothes Washer Manual?

Owner's Manual and Installation Instructions



Three-Speed with Speed Control
AUTOMATIC WASHERS



IMPORTANT:
Read and follow all safety and operating instructions before first use of this product.

Sears, Roebuck and Co., Hoffman Estates, IL 60179 U.S.A. www.sears.com
Sears Canada Inc., Toronto, Ontario, Canada M5B 2B8 www.sears.ca

8526116

PRINTED IN U.S.A.

2/01

ENGLISH ESPAÑOL FRANÇAIS

TROUBLESHOOTING		
Won't Drain/ Won't Spin/ Wet Loads	Drain hose is kinked or clogged.	Straighten or unplug hose (see "Installation Instructions").
	Correct operation of neutral drain.	After a wash or rinse, the washer will drain without spinning. After draining, the spinning begins.
	Lid is open.	Close the lid.
	Drain hose is more than 96 in (2.4 m) above floor.	Lower drain hose (see "Installation Instructions").
	Too much detergent – excessive sudsing.	Use less detergent.
	Wrong cycle for load being washed.	Choose a cycle with a higher spin speed (see "Operating Your Washer").

- Continued search to see if manual turned up; it didn't
- Did a different search specifically for the manual, then looked for info on my problem
- Problem was listed, correct solution (in this case) was not



Writing to Engage with Your Audience

- Previous example underscores how important it is to anticipate users' needs
 - If the information is improperly targeted, is not well-described or is missing, users will not find it
- Know your users!
 - Why they have come to your content?
 - What are they seeking to accomplish?
- This is why having effective personas + scenarios to help guide your technical writers is a priority



How DITA Can Help Shape the Dialogue

- DITA's topic types set the stage for how technical writers communicate with their audience
 - **Concept:** what is this thing and what is it for?
 - **Reference:** what are the correct settings?
 - **Task:** how do I accomplish this procedure?
 - **Troubleshooting:** how do I fix this problem I am having?





DITA + SEO: Summing Up

- Optimize content for search engines and users
- Consider adding a sitemap to help spiders find and index your content
- Descriptive, concise titles/searchtitles!
- Use canonical linking to point to latest-and-greatest topic
- Effective short descriptions can increase CTR
- Relationship tables and the internal links they create also help
- Keep an eye for future developments from Schema.org
- Do not think in terms of SEO “tricks”; best thing you can do is to know your audience and write for them



Further Reading

- Google Webmasters: [google.com/webmasters/](https://www.google.com/webmasters/)
 - Meta tags that Google understands:
support.google.com/webmasters/answer/79812?hl=en
- Sitemaps.org: [sitemaps.org/](https://www.sitemaps.org/)
- The Web Robots Pages: <https://www.robotstxt.org/>
- SEO Pressor Connect Blog: seopressor.com/blog/
- OASIS Feature article on short descriptions (PDF):
oasis-open.org/committees/download.php/57803/
- Zoomin doc references:
 - https://docs.zoominsoftware.io/bundle/zoomin-configuration-guide/page/publish_ss/t_robots.html
 - https://docs.zoominsoftware.io/bundle/zoomin-configuration-guide/page/publish_ss/c_sitemap.html
 - https://docs.zoominsoftware.io/bundle/zoomin-configuration-guide/page/content_manage/t_search_title.html
 - https://docs.zoominsoftware.io/bundle/zoomin-configuration-guide/page/content_manage/c_canonical.html



QA



IXIASOFT

Blog: ixiasoft.com/resources/techcomm-blog-ixiasoft-dita-ccms/

IXIA Talks: ixiasoft.com/ixiasoft-ccms-event-calendar/ixiataalks/

CCMS Link 2019: ixiasoft.com/ixiasoft-ccms-event-calendar/ccms-link-2019/

Twitter: @IXIASOFT (and @KeithIXIASOFT)