

Incorporating Retrievability Into Web Documents

Prepared for InForum '98

Marsha Luevane & Joe Chervenak
National Renewable Energy Laboratory



NREL/PR-24403

Background

- Conducted a two-year project to understand Web search engines and to select a site search engine
- Learned a lot by administering our own site search engine
- Monitored how our documents fare -- and don't fare -- in search engines



Why Optimize?

- Over 175M Web pages are claimed to be indexed
- It's not just trained information professionals who are searching your pages
- Your competitors are optimizing their pages
- You went to all that trouble publishing....
- Lack of controlled vocabulary makes us dependent on keyword searching



Why Optimize? (continued)

- Your Web content deserves the same care , feeding and promotion as
 - Your library catalog
 - Your data bases
 - Your print publications
 - Other document collections



Search Engines 101

- Spider gathers documents
- Search engine parses documents and creates index
- Users search the index



Search Engines 101 (continued)

- Search engines vary in
 - Search algorithms
 - Coverage
 - Update frequency
 - Summary display
 - Ranking method



Optimize in 6 Easy Steps

- Page title
- Meta tags
- Content



Step 1

- Identify important documents
 - Home page
 - Pages linked from home page
 - Main pages of major sections
 - Critical documents



Step 2

- Determine crucial keywords
 - Determine aboutness of important documents
 - Select keywords that express the aboutness
 - Include synonyms

Step 3

- Use descriptive titles
 - Are contained in <HEAD> portion of documents
 - Are used to name Web documents
 - Will appear as document title in search results
 - Should include crucial keywords that describe document content
 - Will boost relevancy ranking in some search engines



Example

<TITLE>Million Solar Roofs Case
Studies</TITLE>

<TITLE>FEMP Project Financing</TITLE>

<TITLE>NREL Photovoltaics Program - What is
PV</TITLE>



Step 4

- Implement description meta tags
 - Are contained in the <HEAD> portion of documents
 - Provide an abstract of document contents
 - Are used as summary displayed in search results
 - Should not exceed 200 characters
 - Will boost relevancy ranking in some search engines



Step 5

- Implement keywords meta tag
 - Are contained in the <HEAD> portion of documents
 - List crucial keywords that describe document contents
 - Should not exceed 800 characters
 - Will boost relevancy ranking in some search engines



Example

```
<HTML>
```

```
<HEAD>
```

```
<TITLE>Department of Energy Photovoltaics Program</TITLE>
```

```
<META name="description" content="PV (photovoltaic or solar electricity) energy technology, how solar cells work, how cells are made into modules and arrays, and common uses of the technology. Photovoltaic news and information, and overview of the national program.">
```

```
<META name="keywords" content="solar energy, solar electricity, solar cells, photovoltaic, solar panels, PV, renewable energy, remote power, power systems">
```

```
</HEAD>
```



NREL/PR-24403

Step 6

- Make the most of your content
 - Content is contained in <BODY> portion of documents
 - Search engines rank keyword frequency
 - Some search engines assign greater weight to terms in beginning of text
 - Position crucial keywords in opening paragraphs
 - Use crucial keywords throughout entire text



Other considerations

- Use meta tags and the ALT attribute in the tag for highly graphical pages
- Use meta tags and the <NOFRAMES> tag to provide information about frame contents
- PDFs have meta data that some search engines index



Tips

- Check search engine results before and after optimization
- Boilerplate approach will burn you
- Beware of deliberate and inadvertent “spamming”



Implementation Strategy

- Who is responsible?
- Who is going to optimize?
- How incorporate into workflow for new documents?
- What about “retro-optimization” of existing documents?



Benefits

- Get your documents to rank better in search engines
- Control how your documents are represented
- Understand your site and documents better and how they stack up against others



Benefits (continued)

- Key part of your strategy to market your site and as well as individual documents
- Makes for good bookmarks
- Creates jobs for catalogers



Resources

- Search Engine Watch
 - searchenginewatch.com
- Search engine Web sites
 - Under “help” or “add URL” sections
- Discussion group
 - news:comp.infosystems.search
- Web Developers Virtual Library
 - www.stars.com

