# Guidelines for Using Quarto Markdown

Ayan Chatterjee, NILU DIGITAL

2024-10-30

## Index

# 1 Quarto Markdown Configuration for Multiple Output Formats

Add the following YAML configuration to the top of the .qmd file to enable multiple output formats, such as html, pdf, and docx:

```
---
title: "Guidelines for Using Quarto Markdown in HTML for Web Crawling"
author: "Your Name"
date: "2024-10-08"
number-sections: true
format:
  html:
    toc: true                # Include a Table of Contents
    toc-title: "Index"
    toc-depth: 3
  pdf:
    toc: true
    toc-depth: 3
  docx:
    toc: true
    toc-depth: 3
sitemap: true                # Enable sitemap generation for web crawlers
keywords: ["SEO", "web crawling", "Quarto Markdown", "HTML"]
description: "This document provides guidelines for using Quarto Markdown in HTML for web cra
---
```

The *toc* option enables a Table of Contents for each specified output type, making navigation easier for longer documents. The *toc-title* option allows to set a custom title for the Table of Contents, which is especially useful for *HTML output*. Additionally, the *toc-depth* option controls the level of headings included in the Table of Contents, allowing to specify how detailed the outline should be, based on the document's heading hierarchy.

The YAML header includes metadata that is critical for *Search Engine Optimization (SEO)* and *web crawling*.

```
---
title: "Optimized Web Crawling Document"
author: "Your Name"
```

```
date: "2024-10-08"
format: html
sitemap: true        # Enable sitemap generation for web crawlers
toc: true            # Include a Table of Contents for better navigation
toc-title: "Index"
toc-depth: 3         # Set TOC depth to include up to h3 headings
number-sections: true
keywords: ["SEO", "web crawling", "Quarto Markdown", "HTML"]
description: "This document provides guidelines for using Quarto Markdown in HTML for web cr
---
```

# 2 Quarto Markdown in HTML for Web Crawling

**Purpose**: This document provides guidelines on using Quarto Markdown to create HTML files optimized for web crawling. These steps and syntaxes will help you structure content, enhance SEO, and improve discoverability of your pages.

## 2.1 Prerequisites

1. **Install RStudio**: Download and install RStudio from [RStudio Download](#).
2. **Install Quarto**: Follow [Quarto installation](#) to install the Quarto CLI.

## 2.2 Basic Setup in RStudio

1. **Create a New Quarto Document**:
   - In RStudio, go to **File > New File > Quarto Document**.
   - Choose the type of document (e.g., HTML) and enter your title and metadata in the YAML header.

2. **Save the File**:
   - Save the file with a `.qmd` extension to ensure it is treated as a Quarto Markdown file.

3. **YAML Header Configuration**:
   - Configure the YAML header with essential metadata to optimize the document for web crawling.

# 3 Essential HTML Structure

## 3.1 Title and Meta Description

Define an appropriate title and meta description in the YAML header, as these are essential for search engines.

```
---
title: "Guide to Quarto Markdown for SEO"
description: "Learn how to use Quarto Markdown to create SEO-optimized HTML content for web
---
```

## 3.2 Headings and Subheadings

Organize content using structured headings (#, ##, ###) to create a hierarchy. This helps crawlers understand the structure and prioritize content.

```
---
# Main Heading

## Subheading 1

Content here.

### Subheading 1.1

More content here.
---
```

## 3.3 Linking Structure

Use descriptive anchor text for links and ensure that internal links are present to improve navigation.

```
---
For more details, refer to the [Introduction to SEO](#introduction-to-seo).
---
```

## 3.4 Image Alt Text and Descriptions

Add meaningful alt text to images to improve accessibility and indexing by search engines.

```
---
![SEO Process](images/seo_process.png){alt="Diagram showing the process of SEO optimization"}
---
```

## 3.5 Tables in Quarto Markdown

Table 1: Table 1: Simple Table Example

| Column 1 | Column 2 | Column 3 |
| --- | --- | --- |
| Row 1, Cell 1 | Row 1, Cell 2 | Row 1, Cell 3 |
| Row 2, Cell 1 | Row 2, Cell 2 | Row 2, Cell 3 |

Table 2: Table 2: Aligned Table

| Left Align | Center Align | Right Align |
| --- | --- | --- |
| Row 1, Cell 1 | Row 1, Cell 2 | Row 1, Cell 3 |
| Row 2, Cell 1 | Row 2, Cell 2 | Row 2, Cell 3 |

Table 3: Table 3: Grid Table Syntax

| Column 1 | Column 2 | Column 3 |
| --- | --- | --- |
| Row 1, Cell 1 | Row 1, Cell 2 | Row 1, Cell 3 |
| Row 2, Cell 1 | Row 2, Cell 2 | Row 2, Cell 3 |

## 3.6 Complex Table with Row and Column Spans

```html
<table>
  <tr>
    <th rowspan="2">Column 1</th>
    <th>Column 2</th>
    <th>Column 3</th>
  </tr>
```

```html
  <tr>
    <td colspan="2">Spanning across 2 columns</td>
  </tr>
</table>
```

# 4 HTML Sitemap Generation for Web Crawling

Enabling the sitemap option in the YAML header creates a sitemap automatically. This sitemap file helps web crawlers discover and index all relevant pages.

## 4.1 Sample Sitemap Configuration

The automatically generated sitemap.xml file might contain entries like the following:

```
---
<url>
    <loc>https://<your-username>.github.io/<your-repo-name>/index.html</loc>
    <lastmod>2024-10-08T12:24:05Z</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.8</priority>
</url>

---
```

## 4.2 Customizing Sitemap

To further customize, use the sitemap: attribute directly in the YAML header to control which pages are included or to add specific pages manually.

## 4.3 Additional Metadata for Social Media and Crawlers

Add Open Graph (og:) and Twitter metadata tags for better social media sharing and visibility.

```
---
meta:
  - name: "twitter:card"
    content: "summary"
  - name: "twitter:title"
```

```
    content: "Guide to Quarto Markdown for SEO"
  - name: "twitter:description"
    content: "This guide helps you create SEO-optimized HTML content using Quarto Markdown."
  - property: "og:title"
    content: "Quarto Markdown for Web Crawling"
  - property: "og:description"
    content: "Optimize HTML content for search engines using Quarto Markdown in RStudio."
---
```

## 5 Quarto Syntax for Key SEO Components

### 5.1 Structured Data with JSON-LD

Use structured data like JSON-LD to help search engines understand the context of your content.

```
---
{
  "@context": "https://schema.org",
  "@type": "Article",
  "headline": "Guide to Quarto Markdown for SEO",
  "datePublished": "2024-10-08",
  "author": {
    "@type": "Person",
    "name": "Your Name"
  },
  "keywords": "SEO, web crawling, Quarto Markdown, HTML"
}
---
```

### 5.2 Linking External Stylesheets and JavaScript

For advanced functionality, link to external CSS and JS files. This enhances the user experience without compromising SEO.

```
---
<link rel="stylesheet" href="https://your-stylesheet-url.css">
<script src="https://your-script-url.js"></script>
---
```

# 6 Rendering the Quarto Document in HTML

Once created the .qmd file, render it in HTML:

## 6.1 Render in RStudio

Go to the RStudio Terminal or Console and run:

```
quarto render yourfile.qmd
```

HTML output: quarto render yourfile.qmd –to html PDF output: quarto render yourfile.qmd –to pdf DOCX output: quarto render yourfile.qmd –to docx

Note: PDF output requires a LaTeX installation. If you haven't installed LaTeX, you can use a lightweight distribution like **TinyTeX** (recommended for R users) or a full installation like **MiKTeX** or **TeX Live**. To install TinyTeX, run:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

We can customize DOCX output with a reference DOCX file by adding reference-doc in the docx configuration.

## 6.2 Preview in Browser

Open the generated HTML file in a browser to ensure the content is well-structured for web crawling and SEO.

# 7 Best Practices Checklist for Accessible and SEO-Optimized Documents

This checklist ensures that your Quarto Markdown documents and tables are optimized for accessibility, SEO, and readability across multiple formats (HTML, PDF, DOCX).

## 7.1 General Document Best Practices

- ☐ **Use descriptive titles and meta descriptions**: Ensure your document has a concise, relevant title and meta description to enhance SEO and search engine visibility.
- ☐ **Include headings with a logical hierarchy**: Organize your document with clear headings (`#`, `##`, `###`) to create a structure that aids readability and search engine indexing.
- ☐ **Add alt text for images**: For accessibility, all images should have meaningful `alt` text describing their content.
- ☐ **Enable the sitemap in the YAML header**: Set `sitemap: true` in the YAML header to improve web crawling and search engine indexing.
- ☐ **Use Open Graph and Twitter metadata for social sharing**: Add `og:` and `twitter:` tags in the YAML header to optimize content for social media platforms.
- ☐ **Include structured data (JSON-LD) for enhanced search engine understanding**: Structured data provides additional context, improving the relevance and accuracy of search engine results.
- ☐ **Link to relevant internal and external pages**: Use descriptive anchor text for links to help guide users and improve SEO.

## 7.2 Accessible Tables Best Practices

- ☐ **Use descriptive headers**: Ensure table headers are clear and descriptive, which aids comprehension and accessibility for screen readers.
- ☐ **Add captions to tables**: Every table should include a caption that summarizes its content to provide context for readers and improve accessibility.
- ☐ **Keep tables simple**: Avoid complex structures with excessive row/column spans unless absolutely necessary for data clarity.
- ☐ **Provide alt text for images within tables**: If images are used within table cells, ensure each image has descriptive `alt` text for accessibility.
- ☐ **Use semantic HTML for complex tables**: For tables with intricate structures, consider using HTML with ARIA roles (e.g., `role="table"`) to improve accessibility and aid screen readers.

Following these guidelines will ensure Quarto Markdown documents and tables are accessible, SEO-optimized, and suitable for multiple output formats.

# 8 Conclusion

This Quarto Markdown document can be saved with a `.qmd` extension, edited in RStudio, and rendered to HTML to ensure it follows best practices for web crawling.