

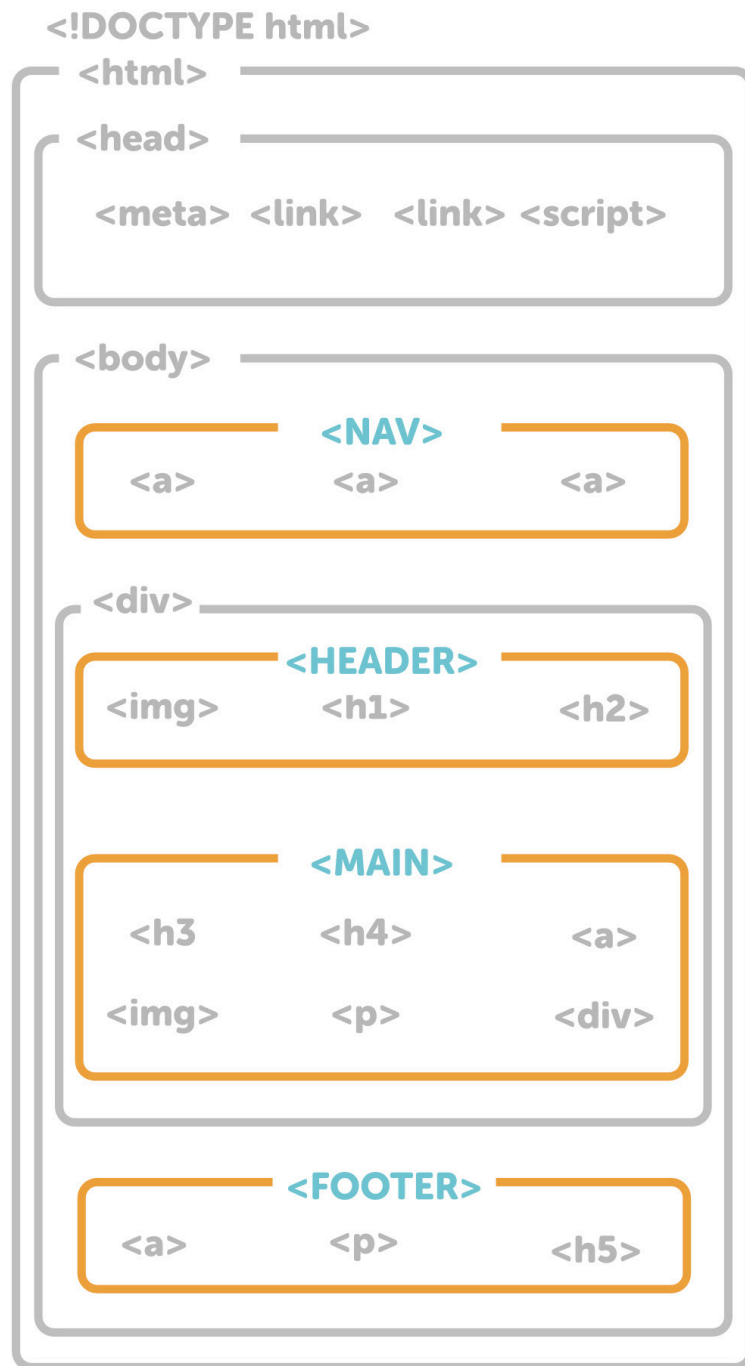
STRUCTURE AND SEMANTICS OF HTML 5

SEMANTICS

Semantics is how we make sense of things.

On the web, when a search engine (e.g., Google) places our site among the top results for a query, it means that it has understood its meaning.

To achieve this, we need to place our linguistic and visual elements into categories that the robots that crawl the web can understand.



```
1  <!DOCTYPE html>
2  <html lang="en">
3  <head>
4      <meta charset="UTF-8">
5      <meta name="viewport"
6          content="width=device-width, initial-scale=1.
7              0">
8      <title>Octavia E. Butler</title>
9  </head>
10 <body>
11     <nav><a href="index.html" target="_self">Home</
12         a> <br />
13     <a href="bio.html" target="_self">Biography</
14         a> <br />
15     <a href="works.html" target="_self">Works</a>
16     <br /></nav>
17
18     <div id=""wrapper">
19         <header>
20             
21             <h1> Octavia E. Butler</h1>
22         </header>
23
24         <main>
25             <p>Welcome to the world of Octavia E.
26                 Butler, a visionary author whose works
27                 have transcended the boundaries of science
28                 fiction.</p>
29             <p>Explore her groundbreaking novels,
30                 short stories, and essays that challenge
31                 societal norms and imagine new futures.</p>
32         </main>
33
34         <footer>
35             <p>2025 Octavia E. Butler. All rights
36                 reserved. <a href="https://octaviaebutler.
37                     com">Visit Official Site</a></p>
38         </footer>
39     </div>
40 </body>
41 </html>
```

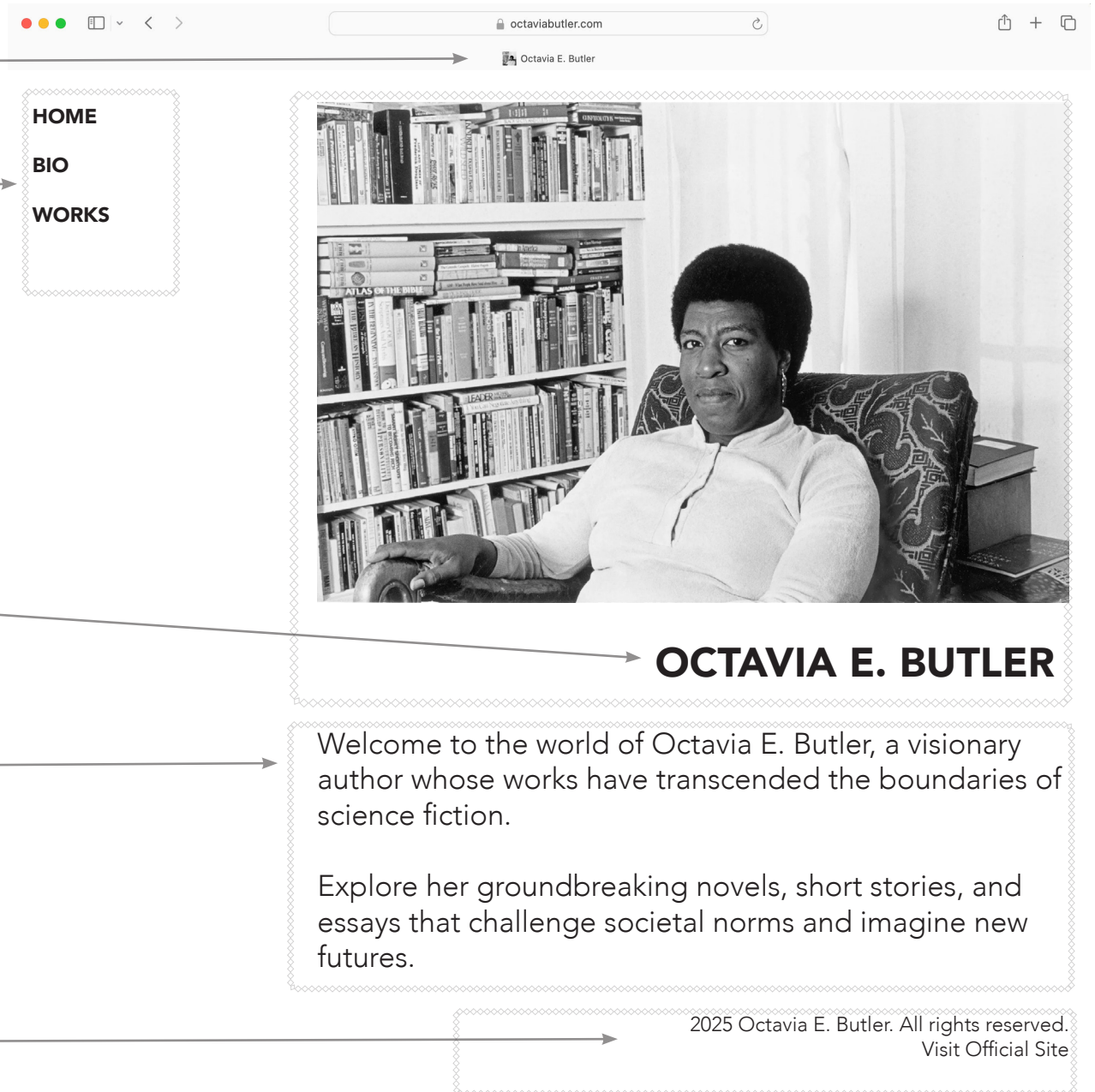
TITLE: Summarizes the content of the page.

NAV: Each of these words is important because it links to another page.

HEADER: The words and images here summarize what is discussed on this page or in this section.

MAIN: Main content of the page. There may be several articles per section and one header per article.

FOOTER: Footer, navigation, and credits.

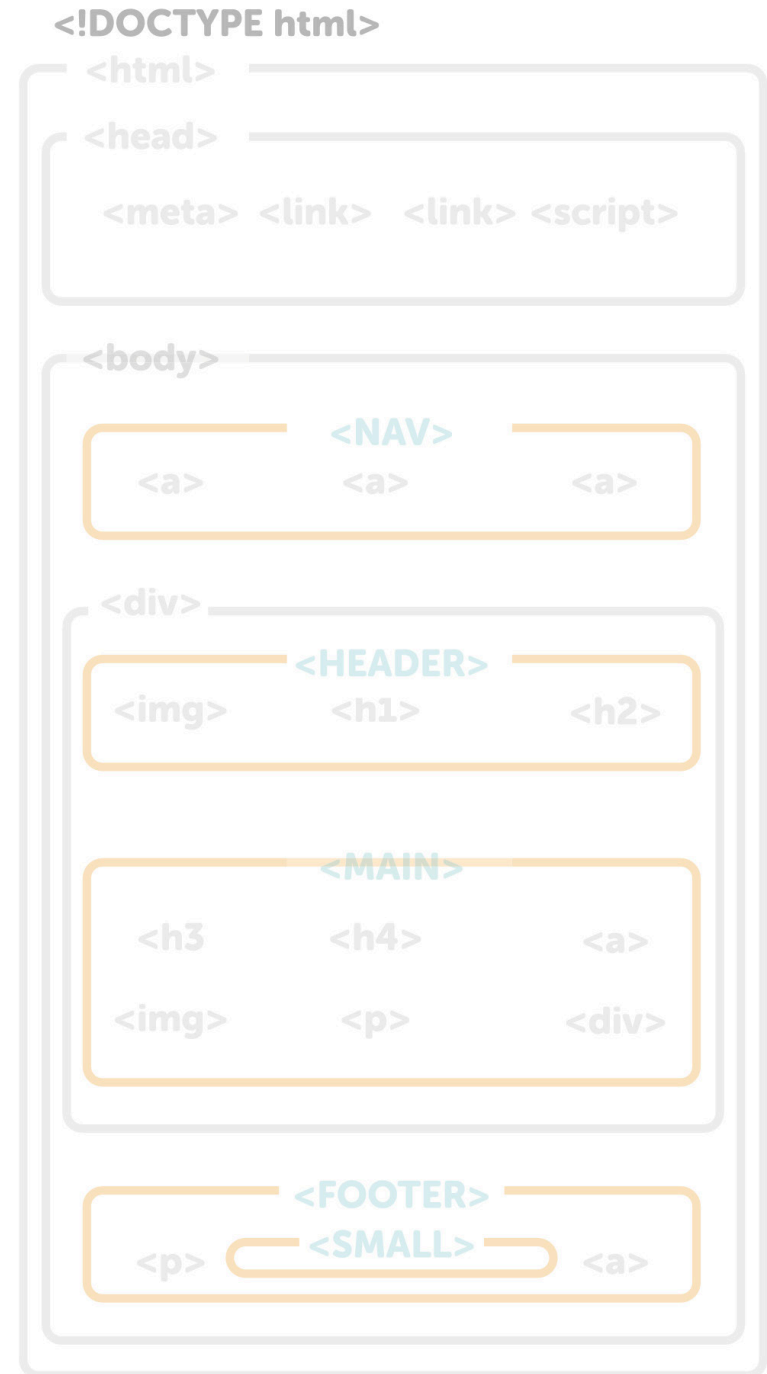


<DOCTYPE>

Informs the browser of the version of HTML that will be used in the page.

Simply add html to specify that it is HTML 5, e.g.:

<!DOCTYPE html>



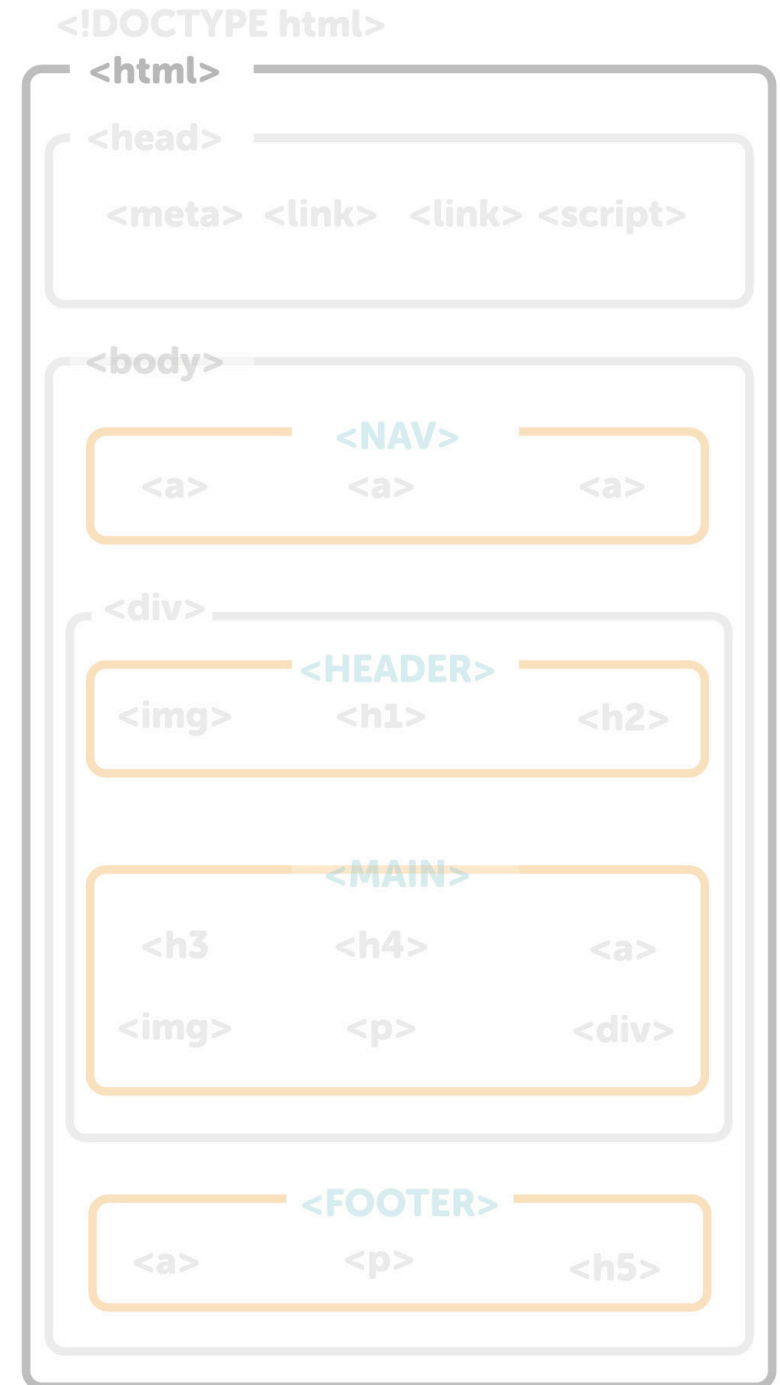
<html>

It defines the beginning of everything that will be on the web page, both for the user and the browser.

There can only be one in the document.

It is a good idea to specify the language of the page using the lang attribute

e.g.: <html lang="en">.

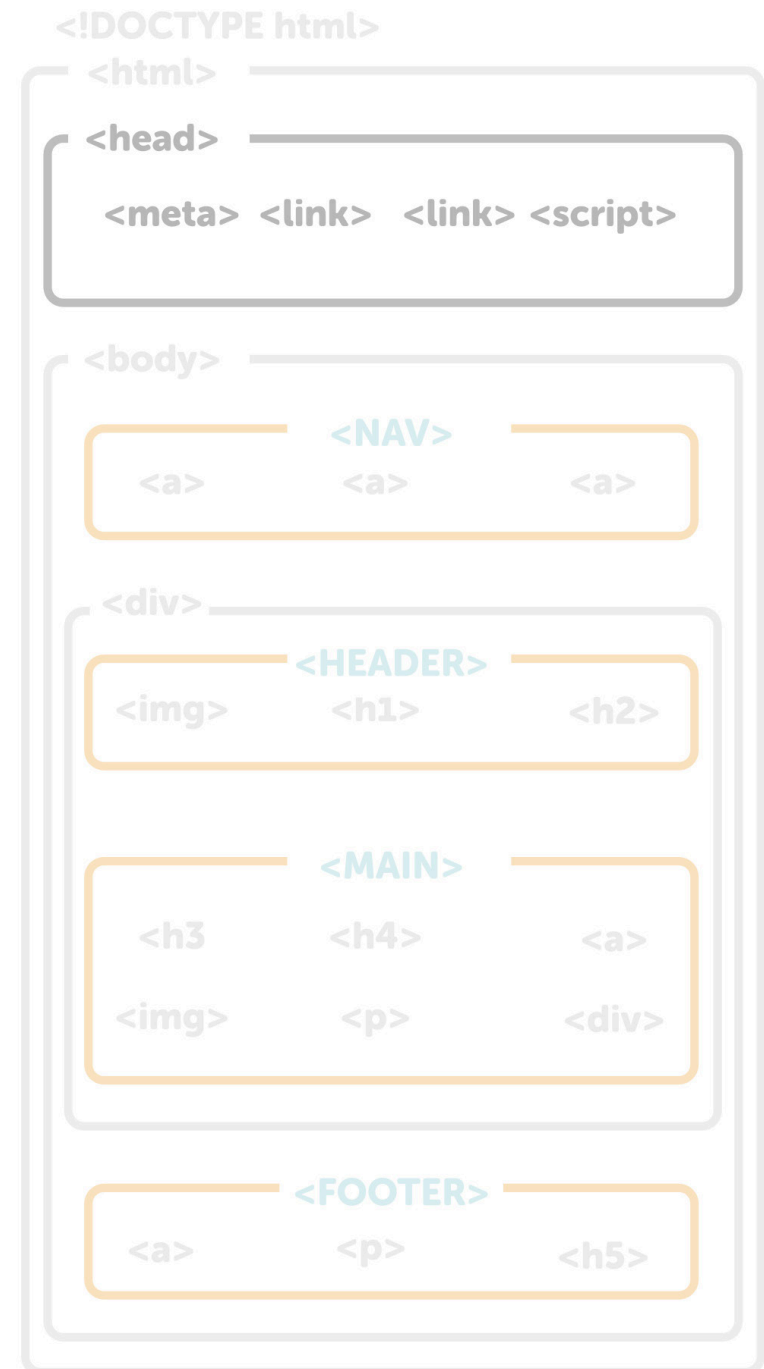


<head>

Invisible to the user, it informs the browser of the page's content with the mandatory <title> tag.

It also informs it of the resources it needs to display the page with the <meta>, <script>, and <link> tags.

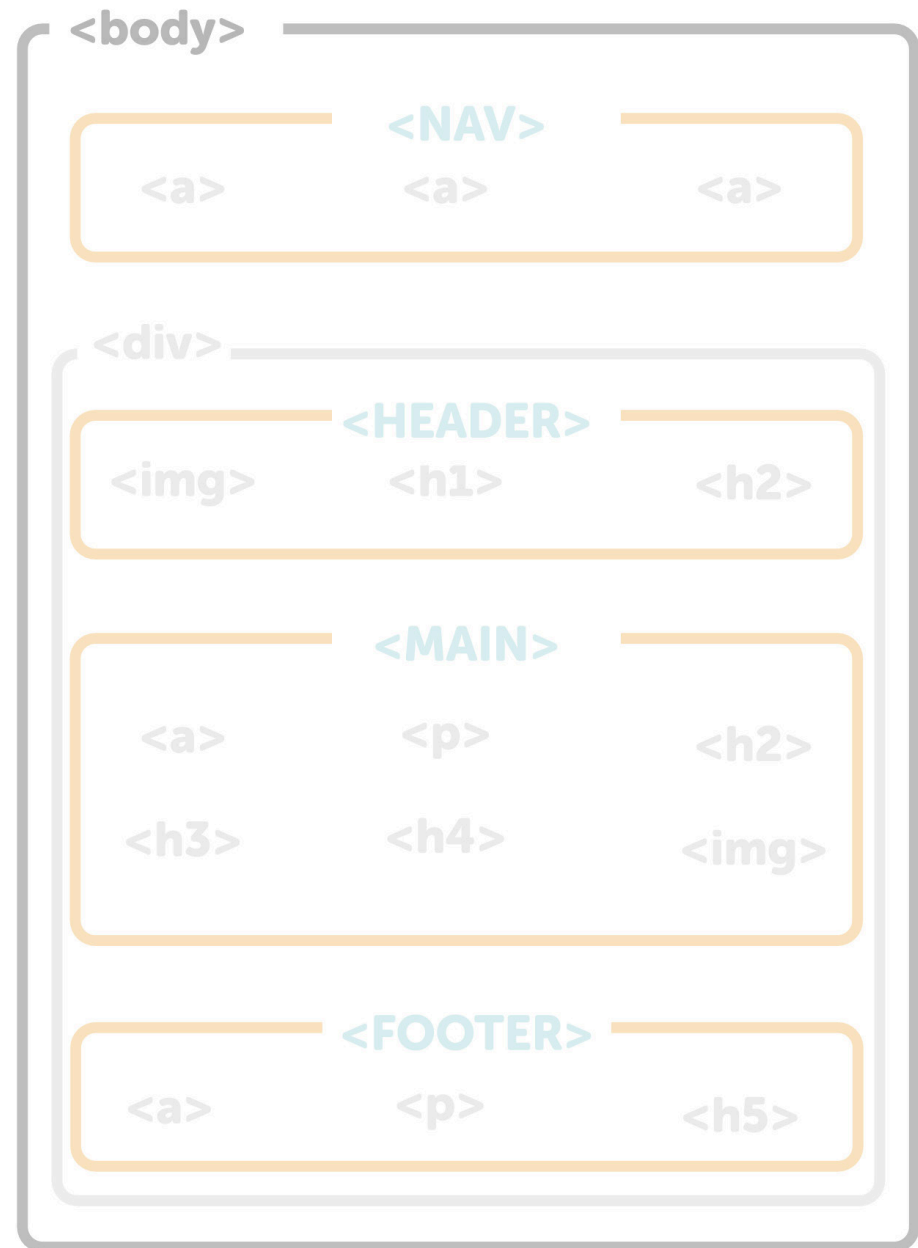
Other meta tags can help index the page with search engines.



<body>

It defines the beginning of what is intended for the user, the page as such.

There can only be one in the document.

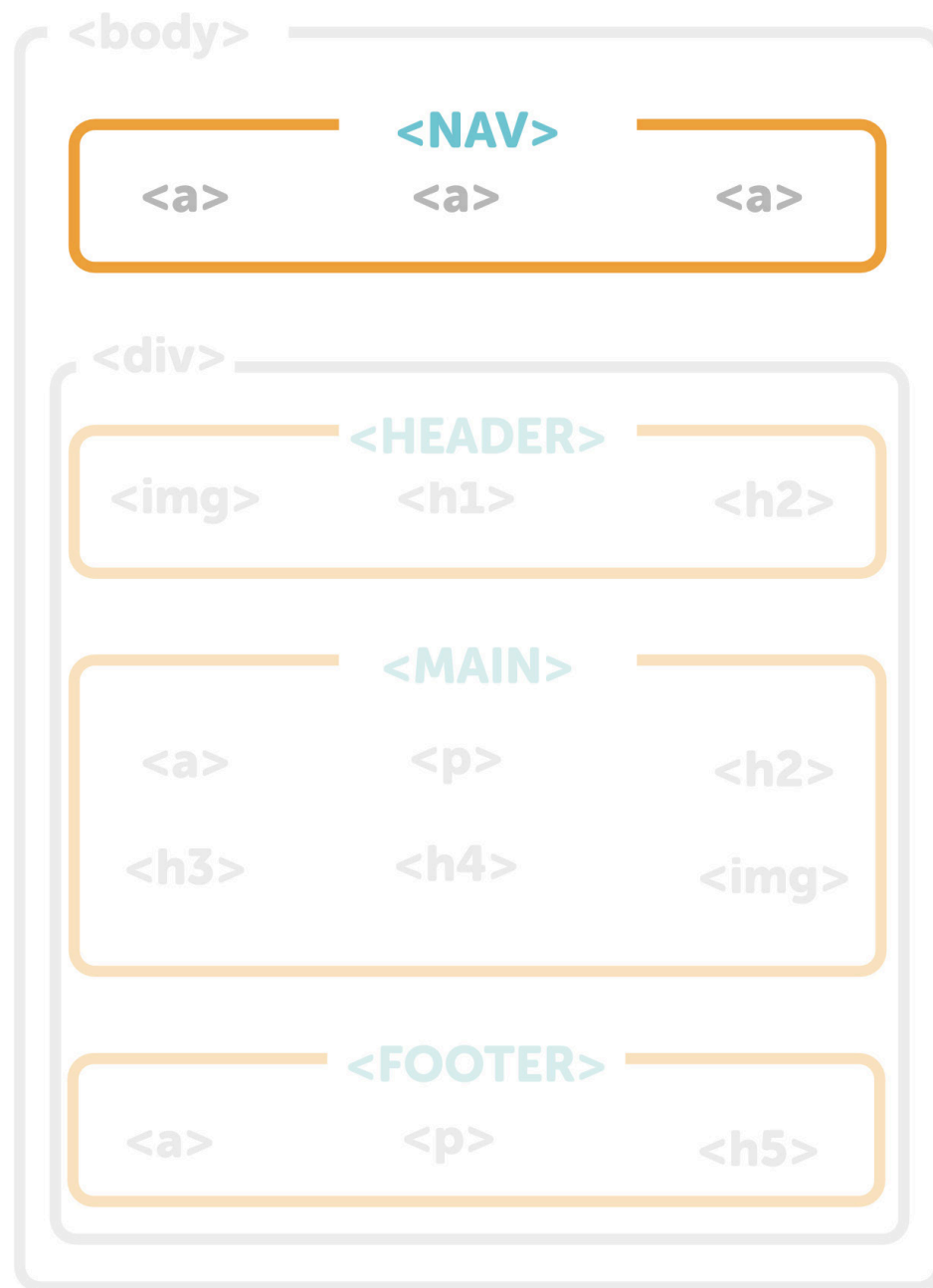


<NAV>

Defines the location of a navigation.

Should only be used for main menus.

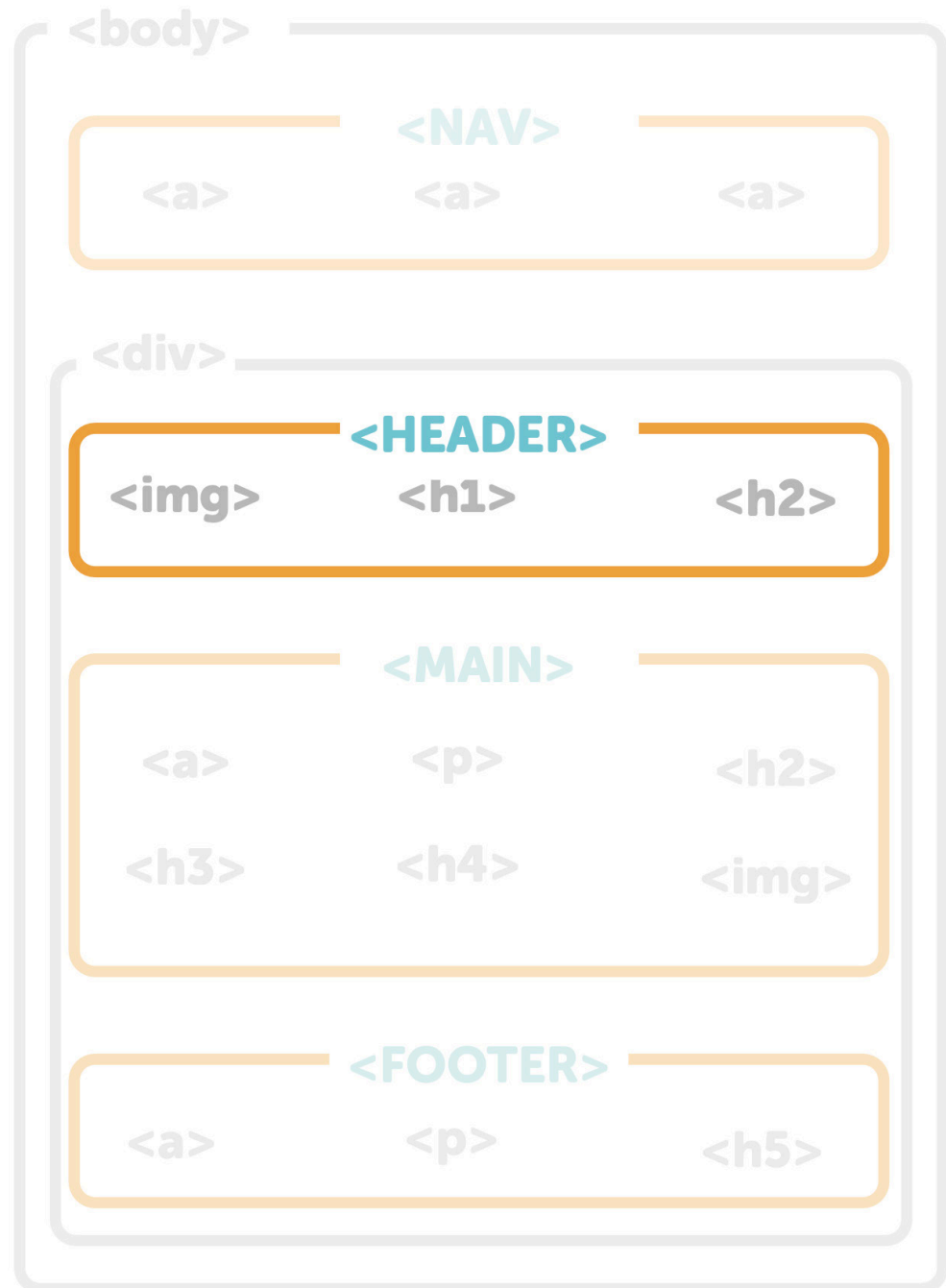
Other <a> links could be placed outside of <nav>.



<HEADER>

Depending on where they are located, they cover the content of a page, an **ARTICLE**, or a **SECTION**.

It is a good idea to put our logo or the title of our site there.



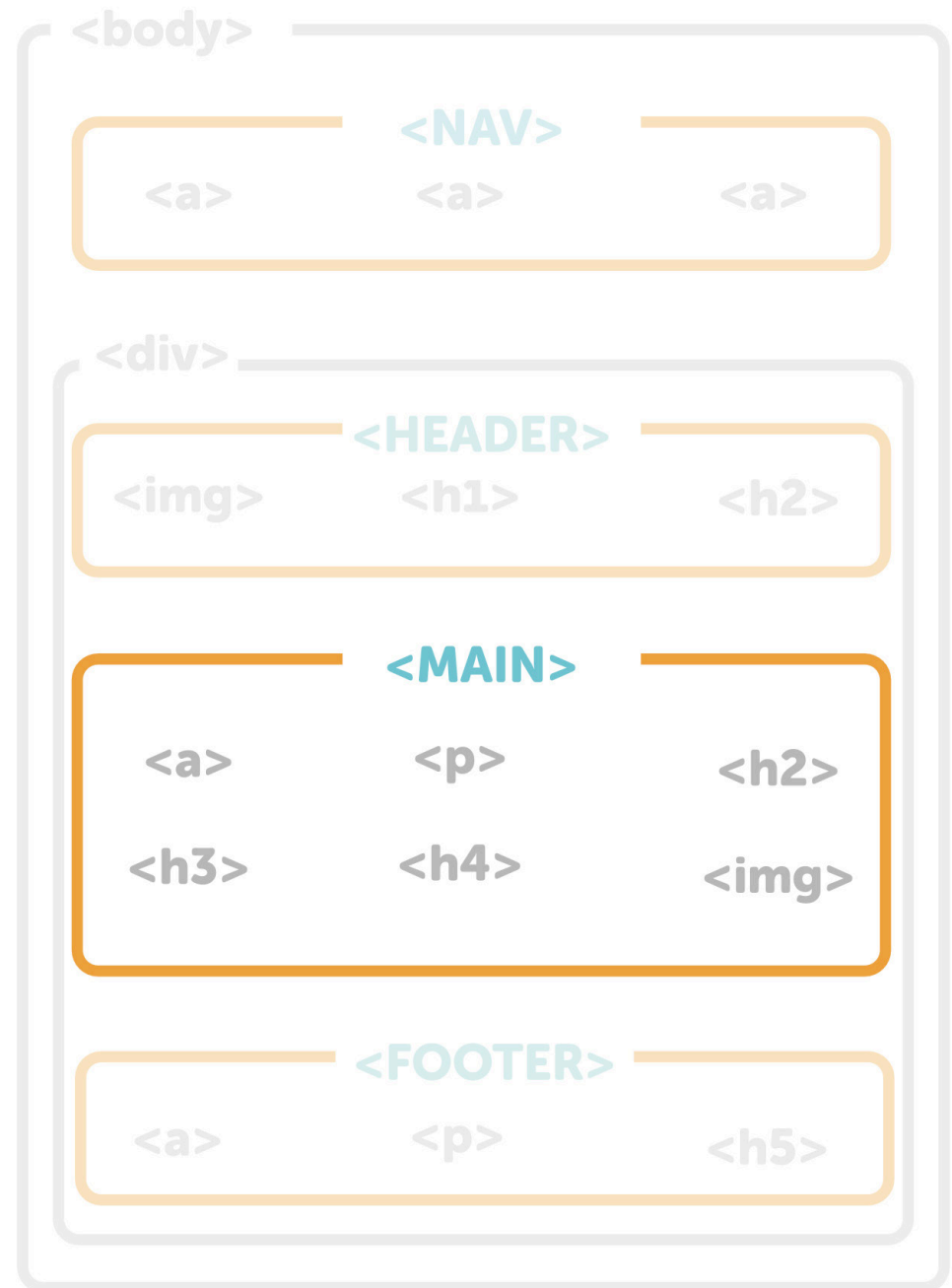
<MAIN>

Defines the main content of the page.

There should only be one per page.

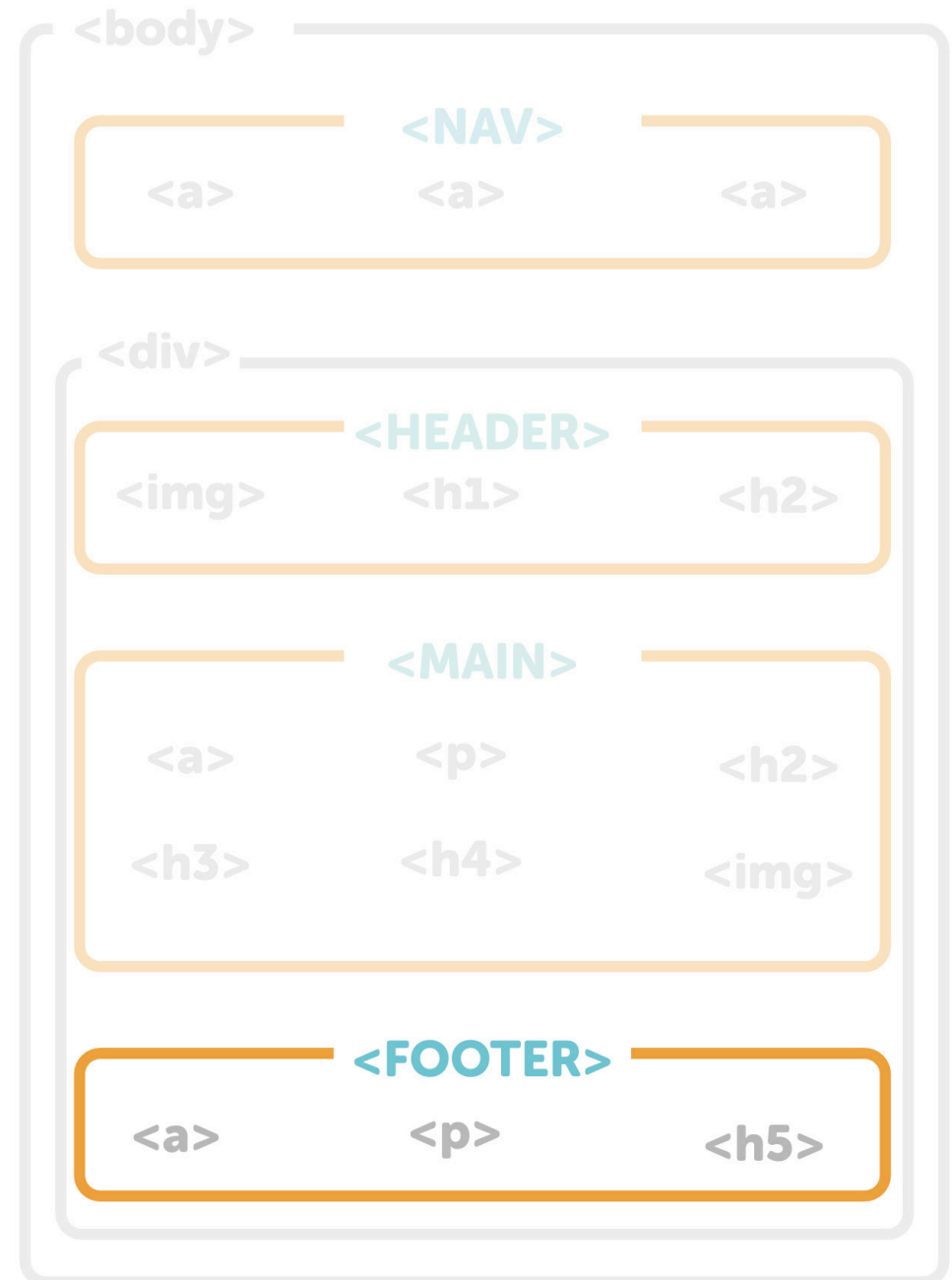
It should not be contained within any other semantic tag.

Elements that are repeated on every page (HEADER, NAV, FOOTER) should be outside of it.



<FOOTER>

**Footer,
where you will find
credits, navigation
reminders, contact
information, and other
relevant and
complementary links.**



<SECTION>

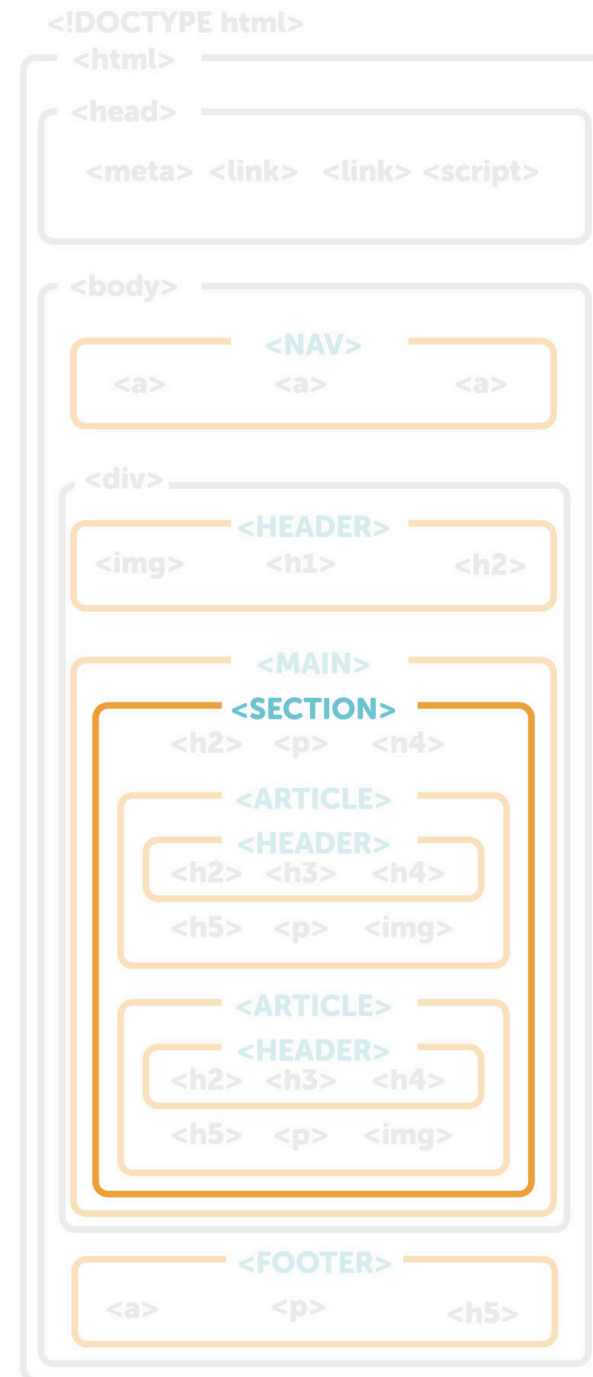
Much like chapters in a book, it would not make sense outside of the page.

Therefore, third-party applications will avoid extracting a **SECTION** from a site on its own.

If headers such as H1, H2, etc. are used within it, it is not necessary to add a **HEADER**.

It can contain several other tags and act as a thematic grouping.

e.g.: contain several **ARTICLES** which will contain **HEADERS** (in this case, it would be like poems in a collection).



<ARTICLE>

Contains elements that together form an independent whole.

An ARTICLE could literally be extracted and appear elsewhere without needing the rest of the page to be fully understood.

e.g.: a blog post, a newspaper article, etc.

It is a good idea to add a HEADER, which will itself contain H1, H2, etc. headings.

It is also acceptable to add a FOOTER at the end of the article if appropriate.

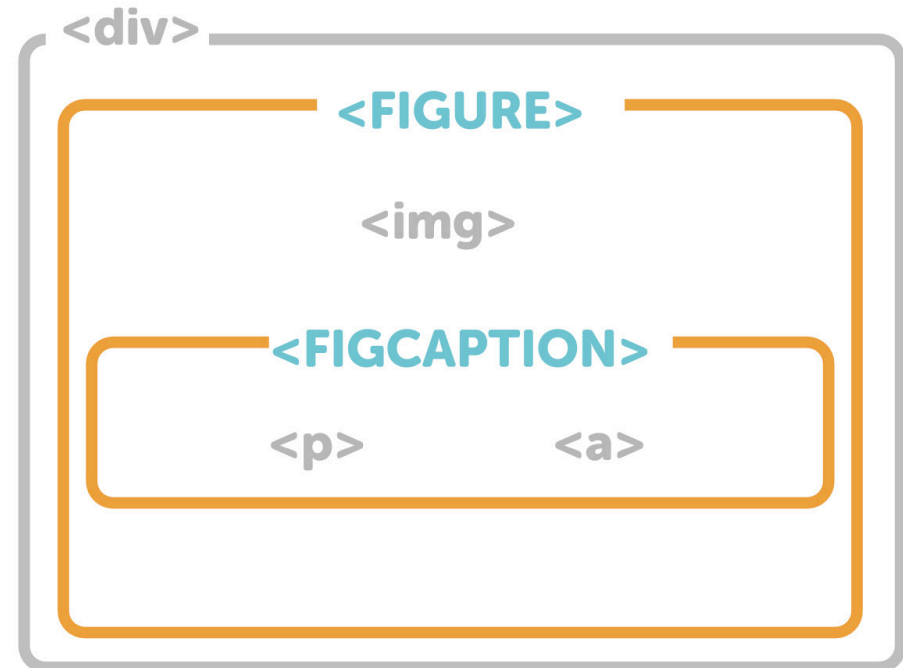


<FIGURE>
<FIGCAPTION>

Figure contains an image, graph, or diagram.

Figcaption must be placed inside Figure and contains the text related to it.

Figure should be able to be extracted from the site and still make sense.



 The text is **bold** without any hierarchical meaning.

 The text is **bold** and is more important than the rest.

<I> The text is *italicized* and is a technical term, comes from another language, is a proper noun, etc.

<Q> The text is *italicized* and is a quotation.

<CITE> The text is *italicized* and represents the title of a book, film, etc.

 The text is *italicized* and means that we want to emphasize it.

 The text should be ~~deleted~~ (will be crossed out).

<INS> The text has been inserted (will be underlined).

<U> The text is underlined and represents a misspelled word.

<SUB> The text is in subscript

<SUP> The text is in superscript