PDF-to-HTML Derivation Algorithm Exercise

Tagging Exercise Generator

January 31, 2026

Contents

1 Document Structure and Sections

This document is intentionally structured to exercise section, article, and division mappings. It includes multiple heading levels, paragraphs, and blocks to drive role mapping.

1.1 Section with an Aside

> This quoted block represents a tangential note that should map to a block quote element in HTML.

Run-in heading (deep level) This paragraph uses a deeper heading level to test mapping of heading levels beyond H6.

1.1.1 Nested Heading Level

This subsection is followed by a short paragraph.

2 Block and Inline Elements

This paragraph includes *emphasis*, **strong text**, and inline `code`. Here is a short inline quote: "quoted content".

Chemical notation: H_2O and $E = mc^2$.

See the external reference and internal link to Section 3.

Este parrafo en espanol debe estar etiquetado con el idioma correcto.

3 Lists

A list starting at 5 with a nested list and then a resumed list:

5. Level A1
    a. Level B1
    b. Level B2
6. Level A2

Some text outside the list, then resume the list:

7. Level A3
8. Level A4

An unordered list:

- First bullet

- Second bullet with *emphasis*

A description list exercising label/body tagging:

Term A Definition for term A.

Term B Definition for term B with an inline `code` fragment.

4 Tables

Table 1: Multi-row header table with header tagging

Quarterly Results		
Department	Q1	Q2
Sales	120	135
Support	98	102
R&D	210	222

5 Figures and Formulas

Inline graphic with ActualText and a decorative artifact: A B

A paragraph containing an inline image should exercise inline figure handling.
Abbreviation example: The mission uses **GPS** (Global Positioning System).
Inline math: $x^2 + y^2 = z^2$.
Displayed math with a matrix:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix}$$

6 Cross References and Links

This sentence references Equation 1 and a section above.
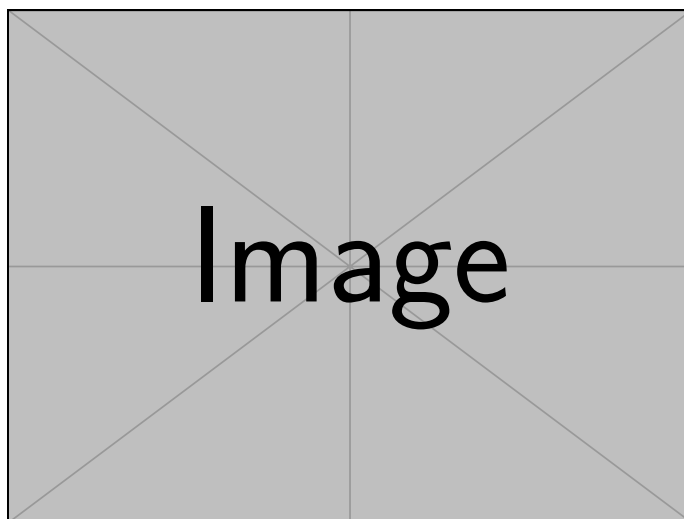
$$E = mc^2 \tag{1}$$

Figure 1: A sample figure with required alternative text

7   Interactive Form Elements

First name:

Email:

Subscribe to updates

Country:

Comments:

8   Bibliography

References

[1] PDF/UA-2 specification, ISO 14289-2.

[2] LaTeX Tagging Project documentation.