# Hygg Test DOCX Document

A comprehensive test document for the Hygg document reader

**Version:** 1.0  
**Date:** 2024-01-01  
**Author:** Hygg Test Suite

## Table of Contents

1. [Introduction](#introduction)
2. [Text Formatting](#text-formatting)
3. [Lists and Structure](#lists-and-structure)
4. [Tables](#tables)
5. [Code and Preformatted Text](#code-and-preformatted-text)
6. [Unicode and Special Characters](#unicode-and-special-characters)
7. [Links and References](#links-and-references)
8. [Block Elements](#block-elements)

## Introduction

Welcome to the Hygg test DOCX document. This file is designed to test various features of document processing in the Hygg reader when working with Microsoft Word format files through pandoc conversion.

This document serves multiple purposes: - Testing text extraction from DOCX files - Verifying formatting preservation - Ensuring Unicode support - Validating pandoc integration

## Text Formatting

This section tests various text formatting options:

**Bold text** is important for emphasis.

*Italic text* provides subtle emphasis.

***Bold and italic*** can be combined.

~~Strikethrough text~~ shows deleted content.

Text can be superscript or subscript (though support varies).

Inline code appears in monospace font.

Here’s a paragraph with **multiple** *formatting* ***styles*** combined in a single ~~line~~ of text.

## Lists and Structure

### Unordered Lists

* First level item
* Another first level item
  + Second level item
  + Another second level
    - Third level item
    - Another third level
* Back to first level

### Ordered Lists

1. First item
2. Second item
   1. Sub-item 2.1
   2. Sub-item 2.2
      1. Sub-sub-item 2.2.1
      2. Sub-sub-item 2.2.2
3. Third item
   1. Sub-item 3.1
   2. Sub-item 3.2

### Definition Lists

Term 1

Definition of term 1

Term 2

Definition of term 2

Can have multiple definitions

### Task Lists

* ☒ Completed task
* ☐ Incomplete task
* ☒ Another completed task
* ☐ Another incomplete task

## Tables

### Simple Table

| Feature | Status | Notes |
| --- | --- | --- |
| Text extraction | Supported | All text content is extracted |
| Formatting | Converted | Formatting is simplified to plain text |
| Tables | Supported | Table structure is preserved |
| Images | Skipped | Images are not included in text output |

### Complex Table with Alignment

| Left Aligned | Center Aligned | Right Aligned | Default |
| --- | --- | --- | --- |
| Row 1, Col 1 | Row 1, Col 2 | Row 1, Col 3 | Row 1, Col 4 |
| This is a longer cell | Centered | Right | Normal |
| Short | Medium length text | 123,456.78 | Last column |

### Table with Spanning Content

| Header 1 | Header 2 | Header 3 |
| --- | --- | --- |
| This is a cell with much longer content that might wrap in narrow displays | Short | Medium |
| A | B | C |
| 1 | 2 | 3 |

## Code and Preformatted Text

### Inline Code

Use the pandoc command to convert files: pandoc input.md -o output.docx

### Code Blocks

def hello\_world():  
 """A simple Python function"""  
 print("Hello, World!")  
 return True  
  
# Call the function  
if \_\_name\_\_ == "\_\_main\_\_":  
 hello\_world()

// JavaScript example  
function factorial(n) {  
 if (n <= 1) return 1;  
 return n \* factorial(n - 1);  
}  
  
console.log(`5! = ${factorial(5)}`);

### Preformatted Text

This is preformatted text  
It preserves spacing  
And line breaks  
 Including indentation

## Unicode and Special Characters

### International Languages

* English: The quick brown fox jumps over the lazy dog.
* Spanish: El rápido zorro marrón salta sobre el perro perezoso.
* French: Le rapide renard brun saute par-dessus le chien paresseux.
* German: Der schnelle braune Fuchs springt über den faulen Hund.
* Russian: Быстрая коричневая лиса прыгает через ленивую собаку.
* Japanese: 素早い茶色のキツネが怠け者の犬を飛び越える。
* Chinese: 敏捷的棕色狐狸跳过懒狗。
* Arabic: الثعلب البني السريع يقفز فوق الكلب الكسول.
* Hebrew: השועל החום המהיר קופץ מעל הכלב העצלן.
* Greek: Η γρήγορη καφέ αλεπού πηδά πάνω από το τεμπέλικο σκυλί.

### Special Characters and Symbols

**Punctuation:** ! @ # $ % ^ & \* ( ) - \_ = + [ ] { }  | ; : ’ ” , . < > / ?

**Quotes:** “Double” ‘Single’ «Guillemets» „German” ‚Single German’ “Curly” ‘Curly’

**Dashes and Hyphens:** - (hyphen) – (en dash) — (em dash)

**Ellipsis:** … (three dots) … (ellipsis character)

**Currency:** $ ¢ £ ¤ ¥ € ₹ ₽ ¥ ₩ ₪ ₨ ₱ ₦ ₡

**Mathematical:** + − × ÷ = ≠ < > ≤ ≥ ± ∓ ∑ ∏ ∫ ∂ ∇ √ ∞ ∈ ∉ ⊂ ⊃ ∩ ∪

**Greek Letters:** α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ τ υ φ χ ψ ω

**Arrows:** ← → ↑ ↓ ↔ ↕ ⇐ ⇒ ⇑ ⇓ ⇔ ⇕ ↖ ↗ ↘ ↙

**Emoji:** 😀 😁 😂 🤣 😃 😄 😅 😆 😉 😊 📚 📖 📝 ✏️ 🖊️ 🖋️

**Other Symbols:** © ® ™ § ¶ † ‡ • ◦ ‣ ✓ ✗ ★ ☆ ♠ ♣ ♥ ♦

## Links and References

### External Links

Visit the [Hygg repository](https://github.com/kruserr/hygg) for more information.

[Pandoc documentation](https://pandoc.org/MANUAL.html) explains format conversions.

### Reference-style Links

The [Hygg project](https://github.com/kruserr/hygg) is a document reader. It uses [pandoc](https://pandoc.org/) for conversions.

### Footnotes

This text has a footnote[[1]](#footnote-44).

Here’s another sentence with a footnote[[2]](#footnote-45).

## Block Elements

### Blockquotes

This is a blockquote. It can span multiple lines and contain other formatting.

It can even have multiple paragraphs.

#### Nested Blockquotes

Level 1 quote > Level 2 nested quote >> Level 3 deeply nested quote > Back to level 2 Back to level 1

### Horizontal Rules

Above the rule

Between rules

Below the rule

### Line Breaks

This line ends with two spaces  
This creates a line break

This is a new paragraph With no line break

### HTML Comments

## Conclusion

This comprehensive test document has demonstrated various features that should be properly handled when converting from DOCX format through pandoc. The Hygg reader should extract all text content while preserving the logical structure and readability of the document.

All formatting should be converted to plain text representation, and all Unicode characters should be preserved correctly.

*End of test document*

1. This is the first footnote. [↑](#footnote-ref-44)
2. This is the second footnote with more content. [↑](#footnote-ref-45)