# 报告标题：Word内容提取功能测试文档

**作者：** Gemini AI

**日期：** 2025 年 12 月 3 日

## 1. 简介与目的

本文档旨在为开发人员提供一个包含丰富格式的测试样本，用于检验文件内容提取工具（如 OCR 或文档解析库）对 Microsoft Word (.docx) 文件中各类元素的处理能力。

## 2. 文本格式测试

### 2.1. 基础样式

这是一个标准的段落。内容应被完整、无误地提取。

**粗体文本**和*斜体文本*是基础的格式。您还应该能够识别出***粗体斜体文本***。

### 2.2. 列表结构

**无序列表 (Unordered List):**

* 第一项：普通文本。
* 第二项：包含 [超链接](https://www.google.com) 的文本。
* 第三项：可以嵌套子列表：
  1. 子列表的第一个元素。
  2. 子列表的第二个元素。

**有序列表 (Ordered List):**

1. 年度总结。
2. 季度计划。
3. 月度报告。

## 3. 表格测试

表格是文档中常见且复杂的结构。提取时应能保持数据的行列关系。

| **区域** | **销售额 (RMB)** | **增长率** | **备注** |
| --- | --- | --- | --- |
| 北区 | 120,000 | +15% | 表现优秀 |
| 南区 | 95,000 | -5% | 需要改进 |
| 东区 | 150,000 | +22% | 领先 |

## 4. 特殊字符与数学公式

测试以下特殊字符和符号的提取：

* 版权符号：©
* 商标符号：™
* 货币符号：¥ € $
* 数学表达式（在 Word 中可能作为嵌入对象或纯文本）：$a^2 + b^2 = c^2$ 和 $\frac{dy}{dx}$.

## 5. 结论

成功从本文档中提取出所有标题、粗体文本、列表结构及其嵌套关系、表格数据和特殊符号，即可认为内容提取功能初步合格。