PDF Rendering Fixes - 2025-11-05

Architecture Diagrams

Issues Fixed

Issue 1: Title Ordering ✅

Problem: Document title appeared BELOW the table of contents

Root Cause: Pandoc generates TOC before processing document body, so H1 titles in the markdown body always appeared after the TOC

Solution Implemented: 1. Modified `generate_pdfs.py` to extract the first H1 title from markdown 2. Remove the H1 from the document body 3. Pass extracted title to Pandoc as metadata: `--metadata title="Document Title"` 4. Pandoc's metadata title automatically renders BEFORE the TOC 5. Removed CSS flexbox hack (didn't work with wkhtmltopdf)

Files Modified: - `generate_pdfs.py`: Added title extraction and Pandoc metadata integration - `pdf-styles.css`: Removed flexbox display and order properties

Result: Title now appears prominently at the top in 28pt bold font, followed by "Index:" TOC

Issue 2: Directory Structure Rendering ✅

Problem: "Project Directory Structure" section displayed as one long run-on line in PDF

Root Cause: Directory tree used box-drawing characters (, ,) which don't render correctly in wkhtmltopdf

Solution Implemented: 1. Ran `fix_all_diagrams.py` to convert directory trees to clean bullet lists 2. The 84-line directory structure with 340 box-drawing characters was condensed 3. Converted to simple bullet list format that renders perfectly in PDF

Files Modified: - `01_system_overview.md`: Directory tree converted from box-drawing to bullet list

Result: Clean, readable directory structure in PDF with proper formatting

---

Verification

Test Case

1. Ran `python3 fix_all_diagrams.py` - Applied fixes to 01_system_overview.md
2. Ran `python3 generate_pdfs.py --file 01_system_overview.md` - Generated PDF with fixes
3. Ran `python3 scan_diagram_issues.py` - Confirmed 0 issues in 01_system_overview.md

Results

- ✅ Title appears before TOC
- ✅ Directory structure renders cleanly
- ✅ No box-drawing artifacts
- ✅ Professional formatting maintained

Technical Details

Title Extraction Algorithm

```python
# Extract first H1 from markdown
for line in lines:
    if line.startswith('# '):
        extracted_title = line[2:].strip()
        # Remove from body
        lines.pop(title_index)
        break

# Pass to Pandoc
pandoc_cmd.extend(['--metadata', f'title={extracted_title}'])
```

Pandoc Command

```bash
pandoc document.md -o document.pdf \
  --pdf-engine wkhtmltopdf \
  --metadata title="Document Title" \   # Title appears BEFORE TOC
  --toc \                                # TOC appears after title
  --toc-depth=3 \
  --css pdf-styles.css
```

CSS Approach (What NOT to do)

```css
/* DON'T DO THIS - Doesn't work with wkhtmltopdf */
body {
    display: flex;
    flex-direction: column;
}
body > h1:first-of-type { order: -2; }
#TOC { order: -1; }
```

Why it failed: wkhtmltopdf has limited CSS flexbox support, and Pandoc's HTML structure makes reordering unreliable

---

Updated Workflow

Correct PDF Generation Workflow

```bash
cd docs/architecture/
```

```
# 1. Generate PlantUML diagrams (if updated)
cd diagrams/
plantuml -tpng *.puml
cd ..

# 2. Resize images to 8 inch max
python3 resize_images.py

# 3. REQUIRED - Fix all text-based diagram issues
python3 fix_all_diagrams.py

# 4. Generate PDFs with proper title ordering
python3 generate_pdfs.py --all

# 5. Verify quality (should show 0 issues)
python3 scan_diagram_issues.py
```

Integration with `/create-architecture-documentation`

The slash command has been updated to include: - Explicit step numbering in the workflow - CRITICAL warning to run fixes before PDF generation - Documentation of all quality fixes applied - Quality assurance verification steps

---

Files Updated

1. `generate_pdfs.py`
   - Added title extraction from first H1
   - Modified `create_temp_markdown()` to return tuple: `(path, title)`
   - Updated `generate_pdf()` to accept title parameter
   - Added Pandoc metadata option: `--metadata title=...`
2. `pdf-styles.css`
   - Removed `display: flex` and `flex-direction: column` from body
   - Removed `order: -2` from h1
   - Removed `order: -1` from #TOC
   - Simplified to standard CSS (no flexbox)
3. `01_system_overview.md`
   - Directory structure converted from box-drawing to bullet list
   - 340 box-drawing characters removed
   - Large 84-line structure condensed
4. `.claude/commands/create-architecture-documentation.md`
   - Added "PDF Rendering Quality Fixes Applied" section
   - Documented all fixes with checkmarks
   - Updated workflow with step numbers and CRITICAL warnings
   - Added quality assurance section

Maintenance Notes

For Future Architecture Documents

1. **ALWAYS** use bullet lists or PlantUML for directory structures
2. **NEVER** use box-drawing characters (, , ,)
3. **ALWAYS** run `fix_all_diagrams.py` before generating PDFs
4. **VERIFY** with `scan_diagram_issues.py` before committing

For Future PDF Improvements

- Title ordering is handled by Pandoc metadata (don't use CSS)

- Text-based diagrams must be converted to proper formats
- wkhtmltopdf has limited CSS3 support (avoid flexbox, grid, etc.)
- Test PDFs after any changes to generation pipeline

Status

Date: 2025-11-05 **Status**: ✔ Complete **Verified**: Yes **Production Ready**: Yes

All PDF rendering issues resolved. System ready for documentation updates.