

What resume parsers actually read

A 45-minute, evidence-first workshop using synthetic PDF and DOCX resumes.

- No participant resumes are needed.
- Every fixture, raw extraction, method, and limitation is public.

One disclosed parser, not a universal compatibility score

The lab runs ResumeGet's production text extractor. It does not certify or reproduce any named applicant tracking system.

- Observe what the release proves.
- Name what it cannot prove.
- Keep negative results visible.

10 layouts. 20 files. 11 risk patterns.

- PDF and DOCX versions of every case
- Headers, footers, columns, tables, symbols, page breaks, decorative spacing, and mixed-language labels
- Reserved synthetic identity, employers, institution, URLs, and achievements

Rendered layout is not extracted reading order

A person sees position, alignment, weight, and proximity. A text extractor returns a sequence.

- Which block is read first?
- Which essential value disappears?
- Which heading becomes separated characters?

Observed results, including the failures

- 14 files preserved every disclosed marker in order.
- 1 file preserved markers but changed reading order.
- 5 files lost at least one disclosed marker.
- 0 files were unreadable by the production extractor.

Essential content in DOCX headers and footers can disappear

In this release, the body extractor omits synthetic identity and contact values stored only as document furniture.

- Keep name and essential contact details in the normal body flow.
- Use repeating furniture for secondary context, not the only copy.

Columns and tables can change sequence

Visual adjacency does not guarantee that extracted education, skills, and experience retain the intended order.

- Inspect the raw text, not only the rendered page.
- Prefer a clear single reading flow for essential information.

One person reads the page. One reads only the text.

- Describe intended hierarchy and actual extraction order.
- Mark missing or moved content.
- Sketch one lower-risk revision and state the visual tradeoff.
- Share the fixture ID, never a participant's personal detail.

Preserve meaning before decoration

- Use ordinary text for section labels.
- Keep essential identity and contact data in body content.
- Use one predictable reading flow.
- Retest after export because PDF and DOCX can differ.

Download the corpus, raw outputs, method, and workshop kit

resume-get.com/research/resume-parsing-lab

- The QR code points to the same canonical page.
- Citation is welcome; a backlink is never required for reuse.