What resume parsers actually read

Workshop contract

- Duration: 45 minutes
- Audience: university career centers, bootcamps, workforce programs, and career educators
- Materials: slide deck, facilitator guide, paired fixture exercise, synthetic corpus, and QR code
- Canonical URL: https://resume-get.com/research/resume-parsing-lab
- Workshop QR URL: https://resume-get.com/research/resume-parsing-lab?utm_source=workshop&utm_medium=offline&utm_campaign=parsing_lab

This session demonstrates the ResumeGet production PDF and DOCX text extractor against a disclosed synthetic corpus. It does not certify or reproduce any named applicant tracking system.

Learning objectives

By the end of the session, participants can:

1. distinguish visual layout from extracted reading order;
2. identify header, footer, column, table, and decorative-heading risks;
3. inspect raw extracted text before drawing a compatibility conclusion;
4. revise a resume layout without submitting a participant's real resume to the workshop facilitator; and
5. explain why a single parser result does not predict every hiring platform.

Timed agenda

Time	Activity
0-5 min	Set the scope, privacy rule, and learning objectives
5-12 min	Compare a rendered resume with its extracted text
12-20 min	Review the disclosed corpus and result labels
20-32 min	Run the paired fixture exercise
32-39 min	Redesign the risky layout and compare tradeoffs
39-45 min	Debrief, share the canonical resources, and collect questions

Facilitator setup

1. Download the workshop bundle before the session.
2. Open the slide deck and keep the facilitator guide available separately.

3. Give each pair one PDF/DOCX fixture and its matching `.txt` extraction.
4. Do not ask participants to send or project a real resume.
5. If participants choose to use the free checker later, direct them to read its privacy notice and make their own decision.

Paired fixture exercise

Roles

- Reader A inspects the rendered synthetic resume.
- Reader B inspects only the matching extracted-text file.

Round one - three minutes

Reader A describes the intended visual hierarchy. Reader B lists the order in which identity, role, experience, education, skills, and languages appear in raw text.

Round two - four minutes

Together, mark:

- any missing expected marker;
- content that moved out of reading order;
- labels whose letter spacing changed extraction;
- content isolated in a header or footer; and
- assumptions that cannot be supported by this evidence.

Round three - five minutes

Sketch one lower-risk revision. Preserve the information, but use a single reading flow, ordinary text labels, and body content for essential contact details. State one visual tradeoff.

Share-out - two minutes

Report the fixture ID, observed label, proposed revision, and one limitation of the test. Do not report a participant name or real resume detail.

Result labels

- Preserved: every disclosed marker is present in the expected order.
- Reordered: every disclosed marker is present, but at least one appears outside the expected order.
- Content loss: extraction completes, but at least one disclosed marker is missing.
- Unreadable: the production extractor rejects the file or returns too little text.

These labels describe a bounded marker check. They do not score resume quality, semantic understanding, visual fidelity, hiring outcomes, or third-party software.

Facilitator notes

On headers and footers

Point out that DOCX body extraction may omit header or footer content even when the rendered file looks complete. Essential identity and contact details should remain in the normal document flow.

On columns and tables

Ask participants to find the first place where extracted order diverges from intended reading order. A visually adjacent block may be serialized before or after unrelated content.

On decorative headings

Expanded letter spacing can turn one word into separated characters. Encourage plain, searchable section labels while allowing decoration elsewhere.

On negative results

Do not repair or hide a failing fixture in the published corpus. A visible failure is useful evidence and makes the limits of the method inspectable.

Privacy guidance

- Use only the included synthetic fixtures during the group exercise.
- Never collect participant resumes, names, email addresses, or checker output.
- Do not photograph projected resume content.
- Do not paste participant content into shared documents or chat.
- Delete any accidental participant submission and stop the exercise until it is resolved.
- The public lab records aggregate, content-free usage under ResumeGet's existing consent boundary.

Accessibility guidance

- Read result labels aloud and do not rely on color alone.
- Share the slide text equivalent before or after the session.
- Describe the resume layout and the extraction-order difference verbally.
- Keep the canonical URL visible as text; the QR code is an additional route, not the only route.
- Allow participants to inspect the files with their own assistive technology.

Closing script

"This lab shows what one disclosed production extraction path observed in a synthetic release. It gives us evidence for improving reading order, not permission to claim universal compatibility. The complete files, raw outputs, method, and limitations are available at resume-get.com/research/resume-parsing-lab."