

SVG Rating Extraction Fix

Branch: `fix/svg-rating-extraction`

Commit: `dd1c1c0`

Date: December 10, 2025

Pull Request: <https://github.com/CliveCaseley/soldcomp-analyser2/pull/new/fix/svg-rating-extraction>

Problem

The EPC scraper was extracting the **wrong rating** from certificates:

- For a certificate showing **"45 E"**, it extracted **"F"** instead
- Root cause: The scraper grabbed the **first A-G letter** found in the SVG
- This was the **scale label** "A B C D E F G", not the actual rating

Analysis

Using debug script (`test-svg-rating-debug.js`), we discovered:

```
SVG text elements:
[0] "A\n B\n C\n D\n E\n F\n G"  ← Scale labels (WRONG)
[6] "45 E"                      ← Actual current rating (CORRECT)
[7] "70 C"                      ← Potential rating
```

The actual rating appears as:

1. SVG element with class `rating-current rating-label`
2. Format: `{score} {letter}` (e.g., "45 E")
3. Also available in `<div class="epc-rating-result">E</div>`

Solution

Implemented 3-tier fallback extraction in `src/utils/epcHandler.js` :

Tier 1: rating-current Class (Most Reliable)

```
$('svg.rating-current, .rating-current').each((i, elem) => {
  const text = $(elem).text().trim().toUpperCase();
  const match = text.match(/\b([A-G])\b/);
  if (match) {
    rating = match[1];
    return false;
  }
});
```

Tier 2: Score+Letter Pattern

```
$('.svg text').each((i, elem) => {
  const text = $(elem).text().trim().toUpperCase();
  const match = text.match(/(\d+)\s+([A-G])\b/);
  if (match) {
    const parentClass = $(elem).parent().attr('class') || '';
    if (!parentClass.includes('potential')) {
      rating = match[2];
      return false;
    }
  }
});
```

Tier 3: epc-rating-result Class

```
$('.epc-rating-result').each((i, elem) => {
  const text = $(elem).text().trim().toUpperCase();
  if (/^[A-G]$/.test(text)) {
    rating = text;
    return false;
  }
});
```

Test Results

```
$ node test-svg-rating-fix.js
```

Certificate	Expected	Extracted	Result
0310-2606-8090-239 9-6161 (45 E)	E	E	✓ PASS
2648-3961-7260-504 3-7964 (64 D)	D	D	✓ PASS

Both tests confirm:

- ✓ Extracts actual rating, not scale labels
- ✓ Uses `rating-current` class successfully
- ✓ Ignores potential ratings

Files Modified

- `src/utils/epcHandler.js` - Fixed `scrapeRatingFromCertificate()` method
- `test-svg-rating-debug.js` - Debug analysis script (new)
- `test-svg-rating-fix.js` - Validation test (new)

PR Status Note

PR #20 was successfully merged!

- Local branch was at commit `0372b4f` (pre-merge)
 - Remote master has commit `e7efb6d` (post-merge with PR #20)
 - When you squash-merge on GitHub, it creates a new commit hash
 - The PR merge was successful - you did everything correctly
-

Next Steps

1. Review the PR: <https://github.com/CliveCaseley/soldcomp-analyser2/pull/new/fix/svg-rating-extraction>
2. Merge when ready
3. This fix ensures accurate EPC ratings for all properties