Information Quality Testing Report

Automated Information Quality Evaluation

1 Context

This report summarizes the assessment of the *Static Accessibility Index Analyzer* using a structured Information Quality (IQ) testing framework. The aim of the analysis was to ensure that the analyzer meets key quality dimensions, focusing on:

- **Free of Error** — verifying that the system produces stable, valid, and deterministic results without unexpected behavior.
- **Interpretability** — ensuring that the outputs are human-readable, transparent, and traceable to their underlying features.

The testing procedure was designed to simulate common issues that might arise during the static analysis process, identify missing or unstable components, and improve the interpretability and robustness of the tool.

2 Testing Criteria

1. Free of Error (FE)

This criterion checks for correctness, consistency, and stability of the analyzer. Each test ensures that the output is reproducible, logically valid, and within expected boundaries.

ID	Description
FE1	**No exceptions:** Confirms that the analyzer completes execution without throwing runtime errors.
FE2	**Unique files:** Ensures that file paths are distinct and there are no duplicate entries in the scan.
FE3	**UTF-8 valid:** Checks that no invalid Unicode characters appear in the scanned files.
FE4	**Deterministic scanning:** Re-runs the file scan to confirm identical hashes, verifying reproducibility.
FE5	**Deterministic detection:** Verifies that detected features are consistent across repeated runs.
FE6	**Score schema valid:** Validates the structure of the score object against a schema using AJV.
FE7	**Score 0–100:** Ensures scoring is bounded within a valid numerical range.
FE8	**No NaN in tags:** Confirms that all tag values are finite and numeric.
FE9	**Stable feature count:** Checks that feature tags exist and are not empty.
FE10	**Baseline consistency:** Ensures that baseline reference outputs remain aligned with expected configurations.

2.1 Interpretability (INT)

This set of tests ensures that the analyzer's outputs are comprehensible and traceable, providing human-readable explanations and transparent scoring mechanisms.

ID	Description
INT1	**Human-readable explanation:** Validates that a clear textual rationale accompanies the computed score.
INT2	**Weights provided:** Confirms that scoring weights are explicitly defined for transparency.
INT3	**Unit specified:** Ensures that the scoring system uses a defined metric (e.g., points).
INT4	**Traceability via counts:** Checks that detailed counts are available for feature traceability.
INT5	**Detections have file and line:** Validates that each detection provides both file name and line number.
INT6	**Threshold config exists:** Verifies the presence of a configuration file (`config.ts`) defining scoring thresholds.
INT7	**Results JSON written:** Ensures that a structured JSON report is generated.
INT8	**README documents scoring logic:** Confirms that scoring methods and metrics are described in documentation.
INT9	**Numeric formatting consistent:** Checks that all counts and summary values are numeric and well-formatted.
INT10	**Interpretability summary complete:** Ensures that the analyzer output includes summaries for high/low pages, feature influence, average index, and grade.

3 Refinements and Adjustments

During the evaluation process, several interpretability tests initially failed, prompting targeted improvements:

3.1 INT6: Threshold Config Exists

The test failed because the analyzer expected a configuration file (`config.ts`) at the project root. To make the test more flexible, the condition was extended to accept the file either in the project root or in the `src/` directory.

3.2 INT10: Interpretability Summary Complete

This test failed because the analyzer's output did not initially include all interpretability elements such as lowest/highest pages or feature influence summary. The analyzer's console and JSON outputs were refined to include:

- Lowest and highest scoring pages for performance visibility.

- An average accessibility index and grade.
- A "Feature Influence Summary" listing the most impactful criteria (e.g., `ariaCount`, `altCount`, etc.).

These refinements improved transparency and interpretability, allowing users to understand not just the score, but the factors influencing it.

3.3 Reflection

The enhanced quality testing framework now verifies both error-free execution and interpretability in the analyzer's results. The refinements to configuration detection (INT6) and interpretability reporting (INT10) have made the system more robust, transparent, and aligned with the goals of information quality assurance.

4 Checks Beyond Automated Testing: Understandability

While most aspects of information quality can be verified through automated or semi-automated tests, certain dimensions such as **Understandability** require human interpretation and cannot be fully assessed by technical checks alone.

4.1 Definition and Relevance

Understandability refers to the degree to which information is clear, easy to comprehend, and effectively communicates its meaning to the intended audience. This dimension is closely related to **Interpretability**, as both aim to improve human comprehension of results. However, understandability extends beyond structure and formatting, it assesses the cognitive effort required to make sense of the information.

4.2 Verification Approach

To verify understandability, automated logic checks (like those used for `INT` criteria) are insufficient. Instead, a **human-centered evaluation** should be carried out. The proposed verification process includes the following steps:

1. **Expert Walkthroughs:** A small group of domain experts is asked to review the analyzer's console output and final report. They are encouraged to perform a "think-aloud" walkthrough, explaining their thought process as they interpret the information.
2. **Structured Questionnaires:** Each participant answers a set of binary (Yes/No) questions designed to measure comprehension, for example:

- Do you understand what the average accessibility index represents?
- Can you tell which accessibility criteria had the most influence on the final grade?
- Do you know how to interpret a "fail" in the test summary?
- Can you link the reported metrics back to specific pages or files?

Each "Yes" indicates that the aspect is understandable, while "No" suggests the need for clearer explanation or visualization.

3. **Scoring and Visualisation:** The results of these binary evaluations are aggregated and visualized using a simple bar or radar chart, where each question corresponds to one dimension of understandability. Higher proportions of "Yes" responses reflect a more comprehensible reporting system.

4.3 Interpretation and Visualisation

The visualisation could use a **radar chart** or **stacked bar graph** where each axis or bar represents a different aspect of understandability:

- Clarity of metrics (e.g., "Average Accessibility Index").
- Traceability of results (e.g., "Linking scores to files").
- Explanation quality (e.g., "Feature Influence Summary").
- Ease of identifying issues (e.g., "Recognizing low-scoring pages").

Participants' binary responses can be converted into percentages to provide a visual summary of perceived understandability. This approach provides a human-centered verification layer to complement the automated interpretability checks already implemented in the analyzer.

4.4 Reflection

While the automated tests effectively measure technical quality and interpretability, they cannot capture how well users actually *understand* the information. By integrating expert walkthroughs and structured user feedback with visualized comprehension data, the analyzer's overall information quality can be validated not only for accuracy and transparency, but also for human usability and cognitive accessibility.