Intersignal Ontology

Version: 1.0\ **Last updated:** 2025-08-10\ **Applies to:** Intersignal Chrome Extension v0.6.x and later

---

1) Purpose

The Intersignal Ontology provides a consistent vocabulary, feature schema, and scoring method for evaluating web pages locally. It enables:

- **Narrative Integrity** scoring (0–100)
- **Diagnostics** for transparency and debugging
- **Ontology Tags** that summarize content traits (e.g., *Thin Sourcing*, *Some Hedging*)
- Stable interfaces between UI, future exporters, and peer agents

Privacy-first: all computation is performed on-device; no page content is transmitted.

---

2) Processing Pipeline (overview)

1. **Acquire** – Extract visible, user-readable text from the active tab; normalize whitespace; respect user selection if present.
2. **Analyze** – Compute feature metrics (hosts, quotes, hedges, title/body overlap, etc.).
3. **Summarize** *(conditional)* – If the browser exposes a local summarizer (e.g., Chrome `window.ai.summarizer`) and the user invokes it, request a short summary; otherwise skip.
4. **Tag** – Map features to Ontology Tags (see §5).
5. **Score** – Compute Integrity Score from weighted features (see §4).
6. **Render** – Display **Integrity Score**, **Summary** (if available), **Diagnostics**, and **Tags**.

---

3) Feature Schema (Diagnostics)

The Diagnostics object is designed to be human-readable and machine-stable.

```
{
  distinctHosts: number,         // unique external hostnames detected
  hosts: string[],               // ordered host list (top 20)
  quotes: number,                // count of quoted substrings (≈ evidence
cites)
  hedges: number,                // count of hedge terms (e.g., "may",
"reportedly")
  titleOverlapTerms: string[],   // stemmed terms overlapping title and body
  titleOverlapRatio: number,     // 0..1 proportion of overlap
```

```
  lengthTokens: number,          // token-length estimate of main body
  linksTotal: number,            // total hyperlinks discovered
  linksExternal: number,         // external hyperlinks (non-self)
  mediaEmbeds: number,           // iframes/players that may carry content
  recencyHints: {
    datesFound: string[],        // ISO-like strings found in body
    relativeTerms: string[]      // e.g., "today", "yesterday"
  }
}
```

Notes

- *distinctHosts* excludes the current page's own domain.
- *quotes* is a proxy for attributable sourcing; not a fact-checker.
- *hedges* is a proxy for epistemic caution; excessive hedging can reduce clarity.
- *titleOverlap* captures headline-to-body alignment (mitigates clickbait).

---

4) Integrity Score (0–100)

Intent: Fast, explainable signal of basic sourcing and alignment. Not a truth oracle.

4.1 Components & Weights (defaults)

- **Sourcing (55%)**
- external links density & diversity (distinctHosts, linksExternal/lengthTokens)
- quotes present (quotes)
- **Alignment (35%)**
- title↔body agreement (titleOverlapRatio, titleOverlapTerms)
- **Clarity (10%)**
- hedge balance (hedges/lengthTokens), extremely low or extremely high hedging both penalize

4.2 Pseudo-formula

```
score = 100 * (
  0.55 * sourcing_index(
           distinctHosts,
           linksExternal,
           lengthTokens,
           quotes)
+ 0.35 * alignment_index(titleOverlapRatio)
+ 0.10 * clarity_index(hedges, lengthTokens)
)
```

Normalization: Each index returns 0..1. Guardrails cap outliers and short texts.

4.3 Interpretation

- **80–100** Strong: multi-source, good quote presence, solid headline alignment.
- **60–79** Moderate: adequate sourcing or alignment; skim with care.
- **40–59** Weak: thin sourcing and/or fuzzy alignment; verify before sharing.
- **0–39** Very weak: likely marketing, speculation, or thin rewrites.

---

5) Ontology Tags

Tags are derived deterministically from features. Multiple tags may appear.

5.1 Sourcing & Structure

- **Multi-source** – `distinctHosts ≥ 3` and `linksExternal/lengthTokens ≥ τ1`
- **Some Sources** – `distinctHosts ∈ {1,2}`
- **Thin Sourcing** – `distinctHosts = 0` or near-zero external links
- **Strong Title Alignment** – `titleOverlapRatio ≥ τ2`
- **Possible Drift** – low `titleOverlapRatio` with frequent tense/tenor shifts
- **AI/ML** – presence of model names, tokens, AI-glossary density
- **Politics / Finance / Health / Science / Culture** – topic cues from n-grams

5.2 Tone & Claims

- **Low Hedging** – `hedges` below lower bound (over-assertive tone)
- **Some Hedging** – within calibrated band (balanced caution)
- **High Hedging** – above upper bound (uncertain/speculative)
- **Quoted Evidence** – `quotes ≥ 2`

5.3 Operational Flags

- **Paywall/Heavy SPA** – excessive script/iframe surface vs text
- **Possible Drift** – re-emerges when title/body semantics diverge
- **Freshness Hints** – recent dates or explicit “updated” cues detected

Thresholds (τ1, τ2, bounds) are tuned per release and documented in the changelog.

---

6) Summary Generation (conditional)

When the browser exposes a local summarizer (e.g., `window.ai.summarizer`):

- **Type:** `key-points`
- **Format:** `markdown`
- **Length:** `short|medium` based on `lengthTokens`
- **Failure modes:** Missing API, pending model download, timeouts on heavy pages.\ Failures never block diagnostics or tags.

7) Examples (non-normative)

1. **Well-sourced explainer**: 6 external hosts, 4 quotes, overlap 0.72 → *Integrity \~86*, Tags: *Multi-source*, *Quoted Evidence*, *Strong Title Alignment*.
2. **Opinion blog**: 0 hosts, 0 quotes, overlap 0.58, high hedges → *Integrity \~42*, Tags: *Thin Sourcing*, *High Hedging*.
3. **Clicky headline, sparse body**: hosts 1, overlap 0.22 → *Integrity \~33*, Tags: *Possible Drift*, *Some Sources*.

---

8) UI Contract

- **Score Ring**: integer 0–100
- **Summary**: string (or banner explaining local model unavailable)
- **Diagnostics**: JSON block per §3
- **Ontology Tags**: ordered badges (max \~5 visible; overflow collapses)

---

9) Privacy & Security

- No content leaves the device.
- No account or telemetry.
- Summarization is local-only; if unavailable, the UI states why and continues.

---

10) Change Control

This ontology is versioned with each release series (0.6.x, R7, etc.). Thresholds and weights may be refined; any changes affecting tag behavior or score distribution are recorded in the **Changelog**.

End of Document