📄 EVIDENCE_SCHEMA.md

Purpose of This Document

This document defines the **formal structure of "Evidence Objects"** used in the system.

An **Evidence Object** represents a **confidence-weighted, extracted claim** derived from user text that can be safely inserted into the **Session Knowledge Graph** and consumed by the **symbolic reasoning engine**.

This schema serves as a **contract between NLP extraction and symbolic reasoning**, ensuring:

- consistency
- auditability
- explainability
- separation of uncertainty from logic

Why an Evidence Schema Is Necessary

Natural language is **uncertain and probabilistic**, while ontology reasoning is **deterministic and symbolic**.

The Evidence Schema provides a **controlled interface** that:

- captures uncertainty **without contaminating reasoning**
- ensures all asserted facts are traceable
- enables causal explanations ("WHY" answers)
- prevents silent assumptions between system modules

Definition: Evidence

> **Evidence** is a structured representation of an extracted mental-health–related concept from user text, annotated with confidence, provenance, and temporal context.

Evidence is:

- ✓ an **input** to reasoning

- ❌ not an inferred conclusion
- ❌ not a diagnosis
- ❌ not a rule outcome

Evidence Lifecycle

```
User Text
   ↓
NLP Extraction
   ↓
Evidence Object (this schema)
   ↓
Session Knowledge Graph
   ↓
SWRL Reasoning
   ↓
Inferred States & Explanations
```

Core Evidence Object Schema

Required Fields

```
{
  "evidence_id": "string",
  "concept_uri": "string",
  "concept_type": "Emotion | Symptom | Trigger | Behavior",
  "confidence": float,
  "source_text": "string",
  "extraction_method": "string",
  "timestamp": "ISO-8601 datetime",
  "session_id": "string"
}
```

Field Definitions

evidence_id

- Unique identifier for this evidence instance

- Used for traceability and explanation provenance

Example:

```
"evidence_id": "ev_2025_09_21_001"
```

---

concept_uri

- Fully qualified ontology URI
- Must exist in `mental_health.owl`

Example:

```
"concept_uri": "mh:Insomnia"
```

---

concept_type

- Ontology class category
- Allowed values only:

```
Emotion
Symptom
Trigger
Behavior
```

---

confidence

- Float in range **[0.0 – 1.0]**
- Represents extractor confidence, not probability of truth

Interpretation guideline:

Range	Meaning
0.80 – 1.00	High confidence
0.60 – 0.79	Moderate confidence
0.40 – 0.59	Low confidence
< 0.40	Not asserted

Evidence with confidence < 0.40 **must not be inserted into the KG**.

---

source_text

- Exact user phrase or sentence fragment that triggered extraction
- Used for explanations and audits

Example:

```
"I haven't slept properly all week"
```

---

extraction_method

- How the concept was extracted
- Allowed values include:

```
keyword_match
pattern_match
semantic_similarity
hybrid
```

---

timestamp

- When the evidence was extracted
- Used for persistence, duration reasoning, and decay

---

session_id

- Links evidence to a specific conversational session
- Enables continuous context reasoning

Optional Metadata Fields

```
{
  "semantic_score": float,
  "matched_pattern": "string",
  "negated": boolean,
  "dataset_reference": "string"
}
```

Optional Field Descriptions

semantic_score

- Raw similarity score (if semantic similarity used)
- Informational only (not used by rules)

---

matched_pattern

- Regex or pattern name that triggered extraction
- Useful for debugging and audits

---

negated

- Indicates negation detected in text

Example:

```
"I am not anxious"
→ negated = true
```

Negated evidence **must not be asserted** into the KG.

---

dataset_reference

- Optional pointer to dataset or study used for validation

Example:

```
"MHP_Figshare_2021"
```

Assertion Rules (Very Important)

An Evidence Object **may be inserted into the Session KG only if**:

✓ confidence ≥ 0.40
✓ negated = false
✓ concept_uri exists in ontology
✓ concept_type matches ontology class

Evidence that fails any condition is **discarded**, not downgraded.

Mapping Evidence to RDF Triples

Each Evidence Object produces **one or more RDF assertions**, e.g.:

```
:SessionUser
   mh:hasSymptom mh:Insomnia ;
   mh:evidenceConfidence "0.82" ;
   mh:assertedAt "2025-09-21T22:15:00" .
```

Evidence metadata may be attached via reification or annotations.

Role Responsibilities

Person 2 (NLP Engineer)

- Produces Evidence Objects exactly in this schema
- Computes confidence values
- Handles negation detection
- Does NOT perform reasoning

Person 3 (Reasoning Engineer)

- Consumes Evidence Objects as inputs
- Aggregates confidence across evidence
- Logs rule firing using evidence_id
- Generates explanations referencing evidence

Person 1 (Ontology Engineer)

- Ensures all concept_uri values exist
- Aligns ontology classes with concept_type
- Maintains semantic consistency

Relation to Explanations (WHY Answers)

All explanations must reference:

- evidence_id
- concept_uri
- source_text

Example explanation fragment:

> "Insomnia was detected (confidence 0.82) from the phrase 'I haven't slept properly all week,' which contributed to AnxietyRisk."

This ensures **full transparency**.

Ethical & Safety Constraints

- Evidence ≠ diagnosis
- Evidence does not imply severity
- Evidence alone never triggers escalation
- Escalation requires explicit safety logic (outside schema)

Why This Schema Strengthens the Project

✓ Explicit uncertainty modeling
✓ Clean NLP–reasoning boundary
✓ Explainability by construction
✓ Audit-ready design
✓ Strong KRR best practice

Final Note

This schema is **authoritative**.

Any module producing or consuming evidence **must comply exactly** with this document.