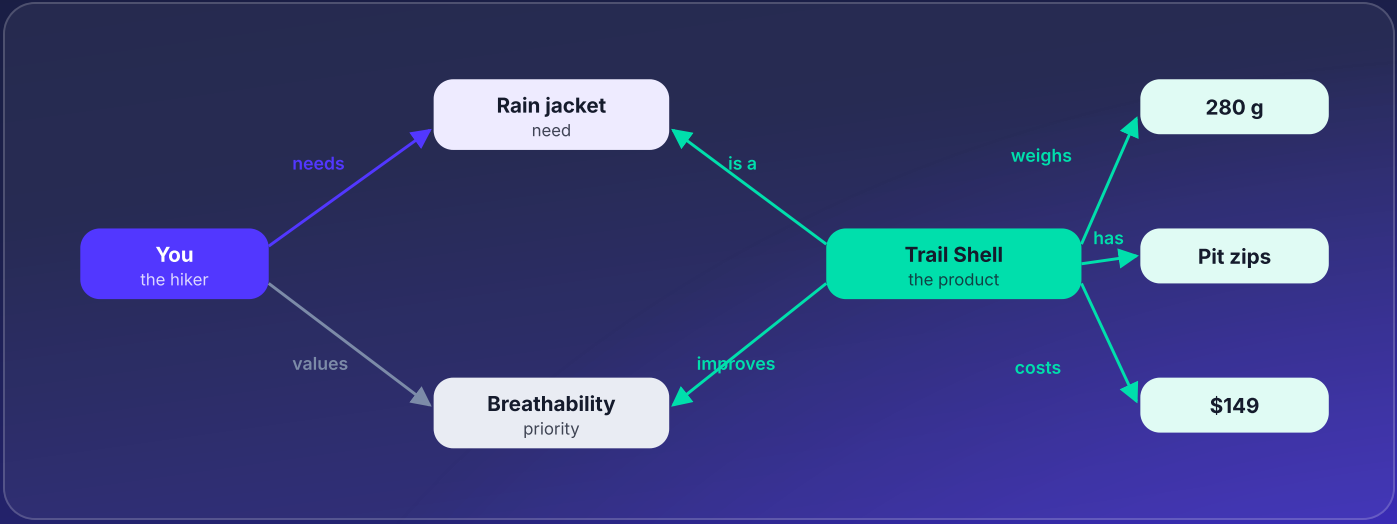
Semantic Triples in Depth

What a triple formally is, how search engines extract and score them, how to choose precise predicates, and how to plan content by joining intent, audience and offer triples into one graph.

Teaching series · Lesson 10 · intermediate to advanced

THREE SETS OF TRIPLES, ONE PARAGRAPH



Search intent · Audience · Offer

6	3	2023
parts	triple sets	notes, corrected

Ameneh Saeednia
Senior SEO and Generative Search Strategist

Before you start

What this lesson covers

Lesson 2 introduced triples: subject, predicate, object, and how to find the main one in a sentence. This lesson goes deeper, using a **2023 training guide from my archive**, corrected and extended with research from Google, the W3C and Microsoft.

By the end of this lesson you can

- Define a triple **precisely**, including what the subject, predicate and object can be.
- Compare triples with **EAV** and **Wikidata statements**, and know when context needs qualifiers.
- Explain how systems **extract, score and trust** triples, based on published research.
- Replace vague predicates with **precise ones**.
- Plan content by joining **intent, audience and offer triples** into one graph.
- **Fact-check** triples before they reach the page or the markup.

Before this lesson

Read **Lesson 2** (triples, graphs and schema) and ideally **Lesson 6** (writing sentences for LLMs). This lesson assumes you can already find the main triple in a sentence.

A note on the 2023 guide

The original guide had a strong method and **several factual and logical errors**. Part 6 shows them openly. Learning to spot them is part of the skill.

Part 1

What a triple formally is

The 2023 guide defined a triple as "three entities that show the relationship between two objects". That is close, but not quite right. **The predicate is a relationship, not an entity, and the object can be a plain value.**

Part	W3C RDF definition [1]	In plain words	Example
Subject	An IRI or a blank node	The thing the fact is about. Usually an identifiable **entity**.	Wayne Gretzky
Predicate	An IRI	The **relationship**. Always a defined property, never free text in a real graph.	member of sports team
Object	An IRI, a **literal** or a blank node	Another entity, **or a value** such as a date, number or word.	Edmonton Oilers (entity) · orange (value)

IRIs make triples unambiguous

An **IRI** is a web identifier, like a URL. "Apple" is ambiguous; a Wikidata IRI or your page's `@id` is not. When your sentence, your markup and Wikidata point at the **same identifier**, every triple lands on one entity.

Literals are where value lives

Most useful facts on a page end in a **literal**: **3 hours, $149, 280 g, ages 6 and up**. Vague copy has no literals, so it produces almost no storable facts.

Remember

Entity → relationship → entity or value. If you can't name the relationship as a short property, the sentence probably isn't stating a fact yet.

Part 2

Triples, EAV and statements

The same fact can be modelled three ways. Knowing the difference helps you decide **how much context a sentence needs**.

Model	Shape	Example	Best for
Triple (RDF)	Subject → predicate → object	Gretzky → member of → Edmonton Oilers	Linking entities across the web
EAV (Lesson 3)	Entity · attribute · value	Trail Shell · weight · 280 g	Product and catalogue data
Wikidata statement	Item · property · value, plus **qualifiers** and **references** [2]	Gretzky · member of · Oilers · *start and end time* · *source*	Facts that are true only in a context

Why qualifiers matter

"Gretzky played for the Edmonton Oilers" is true, but incomplete: he also played for other teams. A statement adds **when** it was true and **who says so**. Your sentences can do the same: add the **date range, location or condition** that keeps the fact true.

Write qualifiers into the sentence

Weak: *Our tour is family friendly.*
Strong: *The tour is **suitable for children aged 6 and up**, and strollers are **allowed on the ground-floor route**.*
The condition is part of the fact.

Rule of thumb

If a fact changes over time or depends on a condition, **state the condition and the date** in the same sentence. A triple without its qualifier can turn into a wrong triple.

Part 3

How machines extract and score triples

Google has published research on turning web text into triples at scale. These papers don't describe today's ranking systems, but they show **how a machine thinks about facts**.

1

Knowledge Vault, 2014

Google researchers extracted facts from the web and fused them with existing knowledge bases. The result held **1.6 billion triples**, but only **271 million** had a confidence of **0.9 or higher**. [3]

2

Knowledge-Based Trust, 2015

A follow-up paper estimated a source's trustworthiness from **the correctness of the facts it states**. It was applied to **2.8 billion facts** and **119 million web pages**. [4]

3

GraphRAG, 2024

Microsoft researchers used an LLM to **derive an entity knowledge graph from documents**, then summarise groups of related entities to answer broad questions. [5]

What this means for writers

Extraction is **probabilistic**. Most extracted facts are not trusted.

Confidence rises when the same fact is **stated clearly, consistently and in more than one source**.

Wrong facts can count **against** a source, not just fail to help it.

Be careful with claims

Knowledge-Based Trust is a **research paper**, not a confirmed ranking factor. Say "Google has researched scoring sources by factual accuracy", not "Google ranks pages by accuracy".

Part 4

Choose the predicate first

The predicate decides whether a sentence states a fact. **Generic verbs** like *is*, *has*, *offers* and *provides* produce weak or misleading triples. Pick the most specific relationship you can defend.

Vague sentence	Weak triple	Precise rewrite	Precise triple
The tour offers a great experience of the old town.	tour → offers → experience	The tour **visits** five sites in the old town.	tour → visits → [5 sites]
Our jacket has great breathability.	jacket → has → breathability	The jacket **vents through** two pit zips.	jacket → vents through → pit zips
The NFL is a type of sports betting.	NFL → is a type of → betting	The NFL **is a** professional American football league.	NFL → instance of → league
The café provides coffee from Ethiopia.	café → provides → coffee	The café **roasts** single-origin beans **sourced from** Ethiopia.	café → roasts → beans · beans → sourced from → Ethiopia

1

Match a real property

Good predicates often match a **schema.org or Wikidata property**: *locatedIn*, *founder*, *duration*, *price*, *weight*.

2

Watch the direction

"Guide **leads** tour" and "tour **is led by** guide" are the same fact. Keep the **main entity as the subject**.

3

Don't hide hedges

"Up to", "from" and "may" change the fact. **"From $149"** and **"$149"** are different literals.

Part 5

Plan content from three triple sets

The best idea in the 2023 guide: before writing, build **three small sets of triples** and join them. The example below uses a rain jacket instead of the original casino example.

1 · Search intent

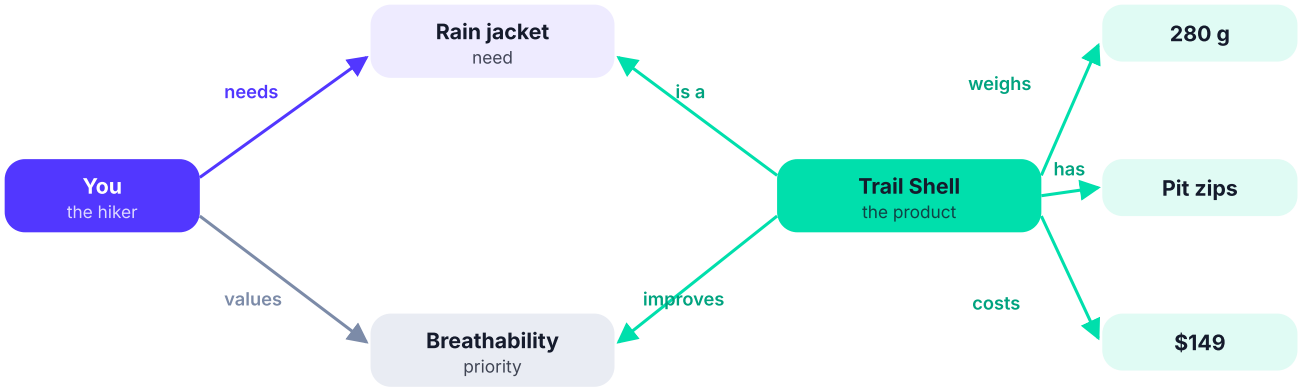
You → need → rain jacket
You → hike in → wet weather
You → want → light gear

2 · Audience

Hikers → value → breathability
Hikers → dislike → heavy packs
Hikers → ask about → packability

3 · Offer

Trail Shell → weighs → 280 g
Trail Shell → has → pit zips
Trail Shell → costs → $149



Where each set comes from

Intent: the query and its follow-up questions (Lesson 6).
Audience: your own reviews, support tickets, surveys or sales calls. **Not** a statistic from a blog post.
Offer: verified product or service facts.

Join on shared entities

The sets connect where they share an entity: the hiker's need (**rain jacket**) and value (**breathability**) meet the product's facts (**pit zips, 280 g**). Every edge is a sentence you can write.

Part 5, continued

From graph to paragraph

The 2023 guide replaced "user" and "persona" with **"you"** before writing. That keeps the copy reader-facing. Then write **one sentence per edge or small group of edges**.

Sentence	Triples it carries
If you hike in wet weather, you need a rain jacket that keeps the rain out without trapping sweat.	You → hike in → wet weather · You → need → rain jacket · Hikers → value → breathability
The [Brand] Trail Shell is a rain jacket that vents through two pit zips, so heat escapes on climbs.	Trail Shell → is a → rain jacket · Trail Shell → has → pit zips · pit zips → improve → breathability
It weighs 280 g and packs into its own pocket, so it won't weigh down a day pack.	Trail Shell → weighs → 280 g · Trail Shell → packs into → own pocket
The Trail Shell costs $149.	Trail Shell → costs → $149

Why this works

The product name is the **subject** of the main triples.
Every claim has a **literal** behind it.
Each sentence still makes sense if an AI lifts it alone (Lesson 6).

What we didn't copy

The 2023 example ended with "Tell your friends!" and promotional filler. **Calls to action belong in the design, not in fact sentences.**

Part 6

Fact-check your triples

A triple that is wrong, or claims more than the evidence supports, is worse than no triple (Lesson 2). Here are the **errors in the 2023 guide**, with fixes.

2023 triple or claim	Problem	Fix
"Welcome to a new age..." becomes A new age → welcomes → online casino gambling	"Welcome" is an **imperative** with an implied subject, "you". The parse was wrong.	Root: (you) → welcome → to a new age. Still no fact about the entity, so **rewrite**.
Persona → likes → progressive jackpots	The source said **only 33%** liked progressive jackpots. A minority became "the persona".	Use the majority finding: *most players preferred **small, frequent payouts***.
Wheel of Fortune → is liked because of → progressive jackpots	**Invented cause.** The source said it was a favourite among the 33%, not why.	Drop it, or state only what was measured.
The user → loves → gambling, so Google matches the page to gamblers	A **speculative chain**. No published evidence shows Google profiles users from triples this way.	Write for the intent you can see. Don't build strategy on guesses.
Survey stats from a LinkedIn post	Secondary source, **no method or sample** shown.	Use **first-party data** or the original study.

The checklist

For every triple: **Is it true? Is it the whole truth (qualifiers)? Who says so? Does the visible page and the markup state the same thing?**

Same triples in markup

Once the facts are right in the text, express the **same triples** in structured data. Each property in JSON-LD is a triple, with the `@id` as the subject.

Product JSON-LD: the offer triples

```
{
  "@context": "https://schema.org",
  "@type": "Product",
  "@id": "https://example.com/jackets/trail-
shell/#product",
  "name": "[Brand] Trail Shell Rain Jacket",
  "brand": {"@type": "Brand", "name": "[Brand]"},
  "category": "Rain jacket",
  "weight": {"@type": "QuantitativeValue",
             "value": 280, "unitCode": "GRM"},
  "additionalProperty": [
    {"@type": "PropertyValue",
     "name": "Packs into own pocket", "value": true},
    {"@type": "PropertyValue",
     "name": "Ventilation", "value": "Pit zips"}
  ],
  "offers": {"@type": "Offer", "price": 149,
             "priceCurrency": "USD"}
}
```

Property	Triple it states
name	#product → name → "[Brand] Trail Shell Rain Jacket"
brand	#product → brand → [Brand]
weight	#product → weight → 280 g
additionalProperty	#product → ventilation → pit zips
offers.price	offer → price → 149 USD

Keep both layers in sync

Google's guidelines say markup must describe the **visible content**. [6] If the page says $149 and the markup says $129, you have stated **two conflicting triples**.

Exercise

Build a page from three triple sets

Time: 30 minutes. Use a real product, service or tour page.

Step	What to do	Output
1	Write **3 to 5 intent triples** from the target query and its follow-up questions.	Intent set
2	Write **3 to 5 audience triples** from first-party evidence. Note the source of each.	Audience set
3	Write **5 to 8 offer triples**, each ending in a **literal** where possible.	Offer set
4	Draw the graph. Circle the **shared entities** where sets connect.	Graph
5	Write one sentence per edge. Replace "user" with **"you"**. Keep the product as the subject.	Draft paragraph
6	Run the **Part 6 checklist** on every triple, then mirror the offer triples in JSON-LD.	Checked copy and markup

What good looks like

Every sentence maps to a triple. No triple **overstates** its evidence. Prices, sizes and times appear **identically** in text and markup.

Discussion

Which audience triple was hardest to support with **first-party evidence**? What would you need to collect to state it confidently?

Check your understanding

Quick quiz

1. In RDF, the predicate of a triple must be:

a) A literal value
b) An IRI that names a relationship
c) A blank node

2. "Tour → costs → $90": the object is:

a) An entity
b) A literal
c) A predicate

3. Wikidata statements add which two things to a basic triple?

a) Qualifiers and references
b) Keywords and links
c) Images and ratings

4. In Knowledge Vault, roughly how many of 1.6 billion triples had confidence of 0.9 or higher?

a) All of them
b) About 271 million
c) About 16 million

5. Which predicate is most precise?

a) offers
b) has
c) is located in

6. A survey says 33% of players like jackpots. The best audience triple is:

a) Persona → likes → jackpots
b) A minority of surveyed players → like → jackpots
c) Players → love → jackpots

Answer key and sources

Answer key

1 b. The predicate is always an IRI. **2 b.** $90 is a value. **3 a.** Qualifiers give context, references give the source. **4 b.** Most extracted facts were not high-confidence. **5 c.** It matches a real property. **6 b.** Don't turn a minority into the persona.

01 **RDF 1.1 Concepts and Abstract Syntax, section 3.1. W3C**
Subject: IRI or blank node. Predicate: IRI. Object: IRI, literal or blank node.
https://www.w3.org/TR/rdf11-concepts/

02 **Wikidata data model. Wikidata Help: Statements**
Statements with properties, values, qualifiers and references.
https://www.wikidata.org/wiki/Help:Statements

03 **Dong et al., Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion. KDD 2014**
1.6B triples; 324M at confidence 0.7 or higher, 271M at 0.9 or higher.
https://research.google/pubs/knowledge-vault-a-web-scale-approach-to-probabilistic-knowledge-fusion/

04 **Dong et al., Knowledge-Based Trust: Estimating the Trustworthiness of Web Sources. 2015**
Applied to 2.8B facts and 119M web pages.
https://arxiv.org/abs/1502.03519

05 **Edge et al., From Local to Global: A Graph RAG Approach to Query-Focused Summarization. 2024**
An LLM derives an entity knowledge graph from documents.
https://arxiv.org/abs/2404.16130

06 **General structured data guidelines. Google Search Central**
Markup must represent the page's visible content.
https://developers.google.com/search/docs/appearance/structured-data/sd-policies

Lesson by Ameneh Saeednia, adapted from a 2023 training guide in my archive, corrected and extended with published research, and prepared with AI assistance. Examples are illustrative.