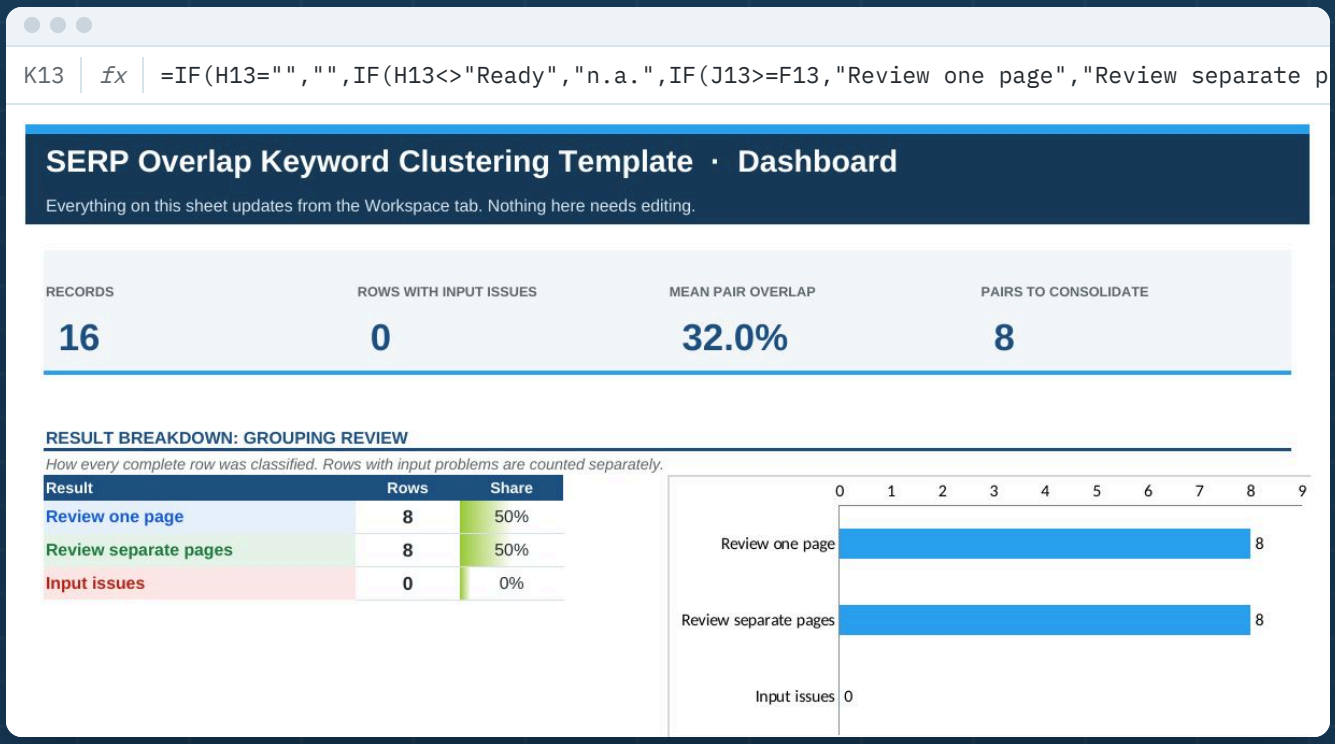
OptimizationTheory.com

Keyword Research Toolkit, resource 03

SERP Overlap Keyword Clustering Template

Decide whether two keywords belong on one page or two by measuring how much their Google results overlap.

Built for the DataForSEO API: SERP API: Google Organic



Jourdan Rombough, optimizationtheory.com

What this template decides

Whether two queries should be targeted by one page or separate pages, based on how many ranking URLs their SERPs share.

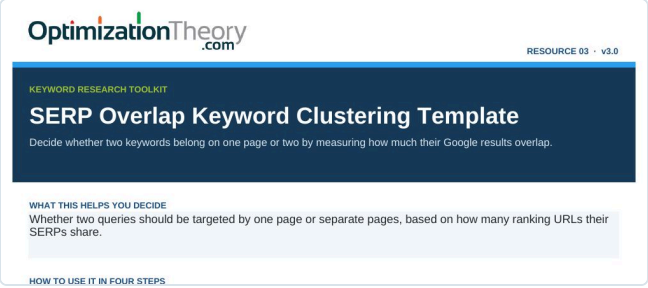
Language models and keyword tools are good at grouping words that look alike. Google doesn't rank pages by how alike the words look; it ranks them by what searchers want. The fastest reliable signal of shared intent is the SERP itself: if the same URLs rank for both queries, Google is treating them as the same need.

This keyword clustering template measures that overlap for each pair of queries. Enter how many URLs you compared for each query and how many appear in both, and it calculates the Jaccard overlap (shared URLs ÷ all unique URLs). Pairs above your cutoff are flagged Review one page; pairs below it, Review separate pages.

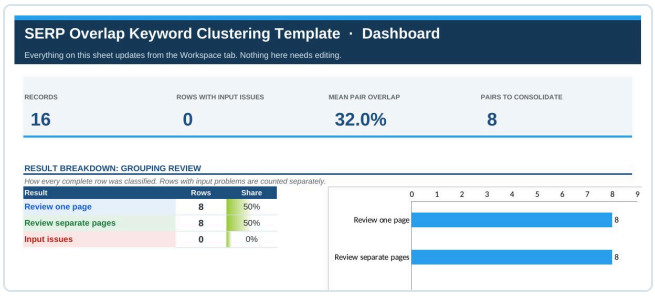
Who it's for

- SEOs clustering a large keyword list into page-level targets
- Writers asking whether a variant needs its own article
- Teams diagnosing cannibalization
- Anyone checking an AI-generated keyword cluster against real SERPs

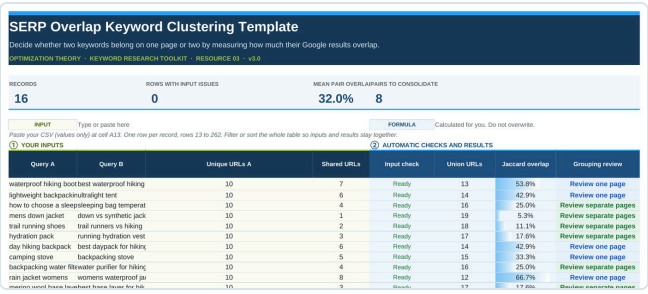
What's in the workbook



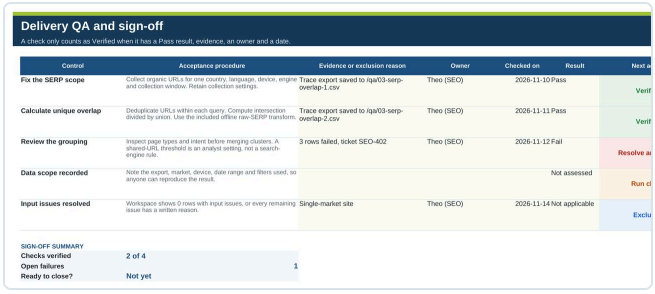
Start Here. Purpose, four-step setup and colour key.



Dashboard. KPIs, charts, Top 10 and an automatic action list.



Workspace. The working table, room for 250 rows.



Delivery QA. Evidence, owner and date for each validation step.

How the calculation works

Union URLs = URLs for A + URLs for B – shared URLs. Jaccard overlap = shared ÷ union. With 10 results each, 3 shared URLs gives 3 ÷ 17 = 0.18, while 7 shared gives 7 ÷ 13 = 0.54. The cutoff is editable per row so you can be stricter for money pages and looser for blog content.

The key formula, as written in the workbook

```
=IF(H13="","",IF(H13<>"Ready","n.a.",IF(J13>=F13,"Review one page","Review separate pages")))
```

Result column	Plain-English formula	What it tells you
Union URLs	Unique URLs A+Unique URLs B-Shared URLs	Unique URLs across both sets.
Jaccard overlap	Shared URLs ÷ Union URLs	Shared URLs divided by union URLs.
Grouping review	IF(Jaccard overlap ≥ Review cutoff,"Review one page","Review separate pages")	Suggestion for human intent review.

The results you'll see

Review one page Review separate pages

Green means the row meets the checks, red means fix it first, amber needs a human review and blue marks an opportunity. Rows with missing or out-of-range inputs show in the Input check column and are left out of the results until fixed.

How to use it, step by step

1

Pick query pairs to test

Start with the pairs you're unsure about: close variants, singular vs plural, 'best X' vs 'X'.

2

Pull the top results

Collect the top 10 organic URLs for each query from the same location, language and device on the same day.

3

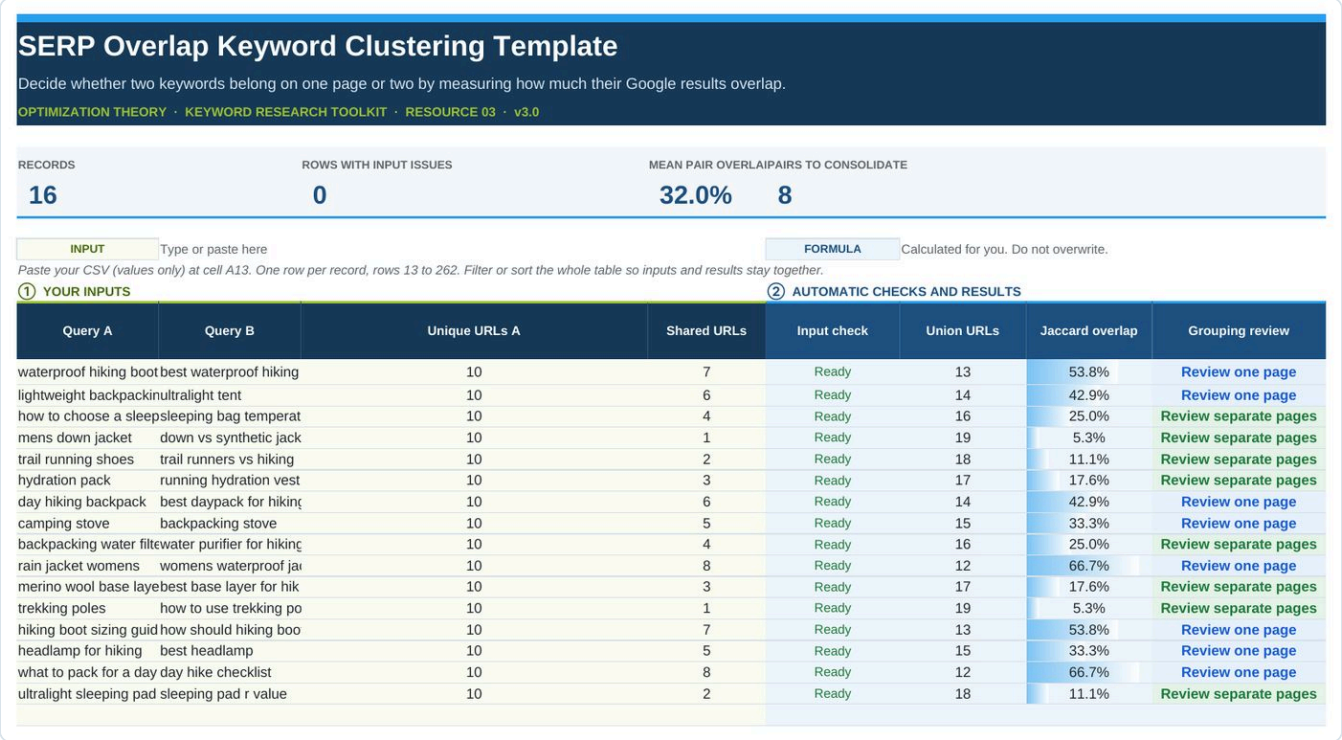
Count shared URLs

Enter the number of URLs compared for each query and how many appear in both lists. Keep the review cutoff at 0.3 unless your niche calls for stricter grouping.

4

Group and map

Pairs marked Review one page go into the same cluster in your keyword map. Pairs marked Review separate pages get their own target URL.



SERP Overlap Keyword Clustering Template

Decide whether two keywords belong on one page or two by measuring how much their Google results overlap.

OPTIMIZATION THEORY · KEYWORD RESEARCH TOOLKIT · RESOURCE 03 · v3.0

RECORDS	ROWS WITH INPUT ISSUES	MEAN PAIR OVERLA	PAIRS TO CONSOLIDATE
16	0	32.0%	8

INPUT Type or paste here — FORMULA Calculated for you. Do not overwrite.

Paste your CSV (values only) at cell A13. One row per record, rows 13 to 262. Filter or sort the whole table so inputs and results stay together.

① YOUR INPUTS — ② AUTOMATIC CHECKS AND RESULTS

Query A	Query B	Unique URLs A	Shared URLs	Input check	Union URLs	Jaccard overlap	Grouping review
waterproof hiking boot	best waterproof hiking	10	7	Ready	13	53.8%	Review one page
lightweight backpackin	ultralight tent	10	6	Ready	14	42.9%	Review one page
how to choose a sleep	sleeping bag temperat	10	4	Ready	16	25.0%	Review separate pages
mens down jacket	down vs synthetic jack	10	1	Ready	19	5.3%	Review separate pages
trail running shoes	trail runners vs hiking	10	2	Ready	18	11.1%	Review separate pages
hydration pack	running hydration vest	10	3	Ready	17	17.6%	Review separate pages
day hiking backpack	best daypack for hiking	10	6	Ready	14	42.9%	Review one page
camping stove	backpacking stove	10	5	Ready	15	33.3%	Review one page
backpacking water filte	water purifier for hiking	10	4	Ready	16	25.0%	Review separate pages
rain jacket womens	womens waterproof jac	10	8	Ready	12	66.7%	Review one page
merino wool base laye	best base layer for hik	10	3	Ready	17	17.6%	Review separate pages
trekking poles	how to use trekking po	10	1	Ready	19	5.3%	Review separate pages
hiking boot sizing guid	how should hiking boo	10	7	Ready	13	53.8%	Review one page
headlamp for hiking	best headlamp	10	5	Ready	15	33.3%	Review one page
what to pack for a day	day hike checklist	10	8	Ready	12	66.7%	Review one page
ultralight sleeping pad	sleeping pad r value	10	2	Ready	18	11.1%	Review separate pages

The Workspace tab with the worked example. Lime cells are inputs; blue cells are formulas.

Fill it with DataForSEO

This template is built to run on **DataForSEO SERP API: Google Organic**: the top organic results for a query.

A helper command builds the input rows from saved API responses. The overlap command compares every pair of saved SERPs from the same location, language and device and fills all six columns for you.

1. Request the data

```
POST https://api.dataforseo.com/v3/serp/google/organic/live/advanced
[
  {
    "keyword": "your keyword",
    "location_code": 2840,
    "language_code": "en",
    "device": "desktop",
    "depth": 10
  }
]
```

Send one task per query with identical location, language and device, and save each response.

2. Save the full JSON response

Save the complete response, not just the task ID, next to the template folder.

3. Convert it to the input CSV

The script in the Automation folder runs offline: no API call, no credentials.

```
python3 Automation/dataforseo_import.py overlap --input serp-*.json --output my-inputs.csv --depth 1
0 --cutoff 0.3
```

Fills automatically	You add
Query A, Query B, Unique URLs A, Unique URLs B, Shared URLs, Review cutoff	Nothing

Source reference fills automatically with the task ID, endpoint, location, language, device and snapshot time, so every row is traceable. Paste the finished CSV at Workspace cell A13. Estimate the cost first with the DataForSEO Cost Calculator (template 38).

Input dictionary

Keep the CSV columns in this order and paste values at Workspace cell A13. A missing observation stays blank; use 0 only for a measured zero or a verified No.

CSV field	Column	Meaning and rule	Typical source
`query_a`	Query A	First query in a matched market and device.	Your records or manual review
`query_b`	Query B	Second query in the same market and device.	Your records or manual review
`urls_a`	Unique URLs A	Distinct organic URLs in the selected depth for query A.	Your records or manual review
`urls_b`	Unique URLs B	Distinct organic URLs at the same depth for query B.	Your records or manual review
`shared_urls`	Shared URLs	Distinct intersection of the two organic URL sets.	Your records or manual review
`review_cutoff`	Review cutoff	Analyst-selected Jaccard cutoff for grouping review. This is not a search-engine threshold. Decimal, 0.08 = 8%.	Your records or manual review
`source_ref`	Source reference	Task ID, export filename and observation scope. Retain the untouched source response separately. Optional.	DataForSEO: source_ref

The worked example

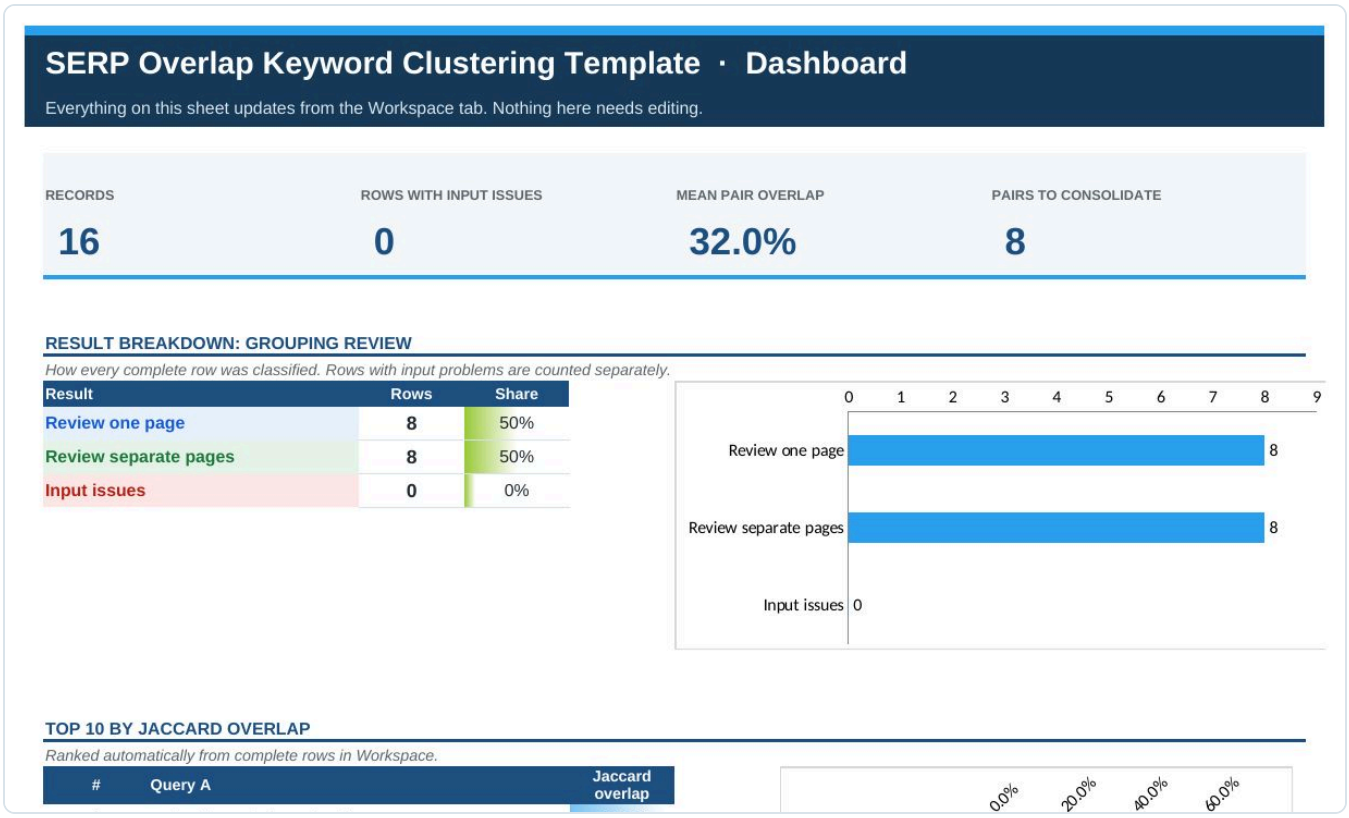
Every workbook ships filled with data from Northbeam Outdoor, a fictional outdoor-gear retailer. This one has 16 example rows.

16	0	32.0%	8
Records	Rows with input issues	Mean pair overlap	Pairs to consolidate

In the example, the boot sizing pair and the rain jacket pair overlap heavily and belong on one page each, while 'trekking poles' and 'how to use trekking poles' share almost nothing: a product page and a how-to guide.

Top 10 by jaccard overlap

#	Query A	Jaccard overlap
1	what to pack for a day hike	66.7%
2	rain jacket womens	66.7%
3	hiking boot sizing guide	53.8%
4	waterproof hiking boots	53.8%
5	day hiking backpack	42.9%



The Dashboard updates automatically from the Workspace tab.

Reading the result honestly

Overlap is supporting evidence for content consolidation. Inspect dominant page types before assigning one URL.

Limits. Pairwise chains can connect dissimilar endpoints. Do not treat connected pairs as an automatically validated cluster.

Mistakes to avoid

Comparing SERPs from different days or locations. SERPs move. Pull both lists at the same time with the same settings or the overlap is noise.

Counting SERP features as URLs. Compare organic URLs only. Videos, shopping and People Also Ask blocks inflate the union and lower the overlap.

Treating 0.3 as a law. The cutoff is a starting point. Check a few borderline pairs by eye and adjust.

Tips from the field

Pull SERPs at scale with the DataForSEO organic endpoint and count overlaps in a script, then paste the counts here.

Re-test clusters for important pages twice a year; intent drifts.

Use the overlap score to settle debates with writers: it's evidence, not opinion.

Questions people ask

What is keyword clustering by SERP similarity?

It groups keywords based on how many of the same URLs rank for them in Google. High overlap means Google sees one intent, so one page should target both.

What overlap means two keywords belong together?

A common rule of thumb is three or more shared URLs in the top 10, roughly a Jaccard score of 0.18 to 0.3. This template defaults to 0.3 for a stricter grouping.

What is Jaccard similarity?

It's the size of the intersection of two sets divided by the size of their union. Here, shared ranking URLs divided by all unique ranking URLs across both queries.

Can I use this with Google Sheets?

Yes, the XLSX opens in Google Sheets and Excel with all formulas working.

Templates that work well with this one

Keyword Mapping Template
Assign every keyword to exactly one URL and catch conflicting assignments before they become cannibalization.

Keyword Cannibalization Checker Template
Find queries splitting clicks across two URLs, check whether intent really overlaps and decide whether to consolidate.

Keyword Prioritization Template
Rank every keyword by the profit it could add per hour of work, not by search volume alone.

Keyword Gap Analysis Template
Find the queries your competitors rank for that you don't, and size each gap in potential clicks.

SEO Content Brief Template
A content brief built for Google and AI answers: real questions, original evidence, sources and a three-step approval workflow.

AI Prompt Research Template
Turn real customer tasks and keywords into a weighted prompt panel for ChatGPT, Perplexity, Gemini and AI Overviews.

Read the full tutorial online: **optimizationtheory.com/download-view/keyword-clustering-template/**
Get all six templates in the **Keyword Research Toolkit**, or all 40 in the **Complete SEO & GEO Toolkit**.