

Content Pipeline

Automatizovaný systém pro sběr, hodnocení, generování
a publikaci obsahu

Jan Kuneš | Case Study | v1.0 | Březen 2026

Výzva

Manuální sběr

Redaktor denně prochází desítky zdrojů a hledá relevantní témata

Nekonzistentní kvalita

Bez systémového hodnocení se volba témat řídí intuicí

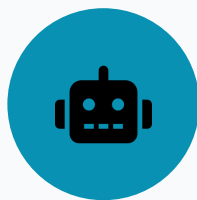
Časová náročnost

Od nápadu k publikaci trvá hodiny až dny

Škálování

Provoz více obsahových webů vyžaduje násobný tým

Řešení: 5-krokový pipeline



Sběr

RSS feedy

Hodnocení

AI scoring

Témata

AI návrhy

Články

AI generování

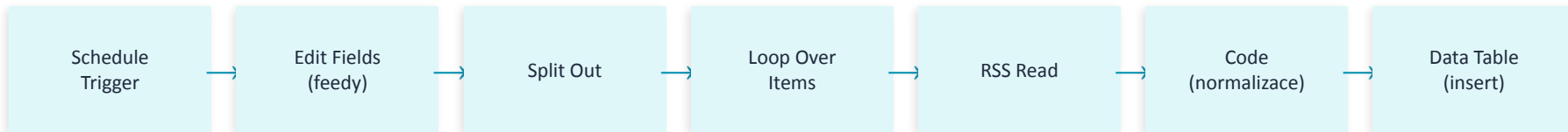
Publikace

WordPress

Každý krok je samostatné n8n workflow. Komunikují přes sdílené Data Tables.

A Feed Collector

Účel: Automatický sběr článků z RSS feedů 2× denně



Klíčové vlastnosti:

- 5 RSS zdrojů (GSMArena, Android Authority, 9to5Google, 9to5Mac, The Verge)
- Deduplikace přes If Row Does Not Exist — existující články se nepřepisují
- Automatický freshness_score (0–100) podle stáří článku
- Source mapping přes regex (new URL() není v n8n sandboxu)

B Source Enrichment & Scoring

Účel: AI hodnocení atraktivity, relevance a potenciálu každého článku

Scoring model:

Freshness 25%	Audience fit 20%	Readability 15%	Novelty 15%	Commentary 15%	Source priority 10%
------------------	---------------------	--------------------	----------------	-------------------	------------------------

external_interest_score ≥ 60 \rightarrow candidate | < 60 \rightarrow discarded

AI klasifikace (GPT-5.4-mini):

- main_topic, subtopics, entities
- audience_fit_score, readability_score, novelty_score
- commentary_potential, practical_value, possible_angles

D Topic Generator

Účel: Z kandidátů generovat originální témata pro vlastní web

Proces

- Top 30 kandidátů podle skóre
- Všichni najednou do GPT-5.4
- AI navrhne 5–10 originálních témat
- Deterministický final_score
- ≥ 65 = proposed | < 65 = backlog

Výstup pro každé téma

- proposed_title (v češtině)
- angle — unikátní úhel pohledu
- why_now — proč právě teď
- why_us — proč právě my
- formát (analysis, guide, opinion...)

F Article Generator

Účel: Generování celých článků (1000–2000 slov) v češtině

Pravidla generování

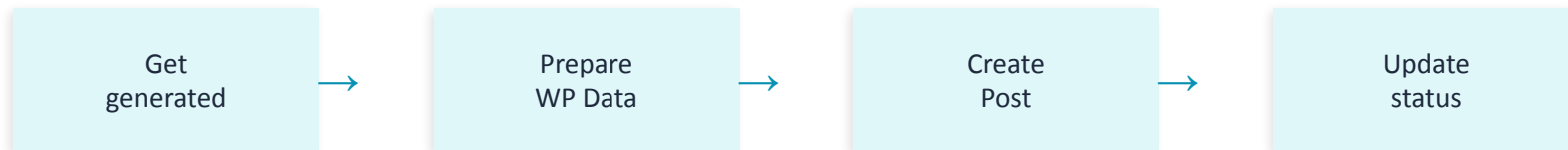
- Publicistický styl, přirozená čeština
- Titulek, perex, úvod, 3–5 sekcí, závěr
- Ceny v Kč, dostupnost v ČR
- Názorový, ale podložený fakty
- Krátké odstavce (max 3–4 věty)
- Ruční schválení tématu (approved)

Výstup

- `final_title`
- `perex` (max 160 znaků)
- `article_body` (Markdown)
- `meta_description` (SEO)
- `tags`
- `word_count`

G WordPress Publisher

Účel: Automatická publikace článků na WordPress včetně excerpt



- Markdown → HTML konverze (H2, H3, bold, odstavce)
- Excerpt + title + content v jednom HTTP Request
- Automatický status update na „published“
- WordPress REST API s Basic Auth

Datový model

Tabulka	Sloupců	Účel
source_articles_raw	14	Surová data z RSS feedů
source_articles_enriched	14	AI hodnocení externích článků
topic_candidates	12	Návrhy témat pro vlastní web
generated_articles	13	Vygenerované články

Tok statusů

- raw: new → candidate / discarded
- enriched: candidate → topic_generated
- topics: proposed → approved (ručně) → generated
- articles: generated → published

Lokalizace pro český trh

Hodnocení (B)

Produkty dostupné v EU/ČR
Evropské ceny a regulace
Penalizace US-only obsahu

Témata (D)

Titulky v češtině
Ceny v Kč
Samsung, Apple, Xiaomi,
Poco, Motorola, Nothing

Články (F)

Publicistická čeština
Ceny v Kč
Dostupnost v ČR
Žádný překladový jazyk

Výsledky & další kroky

Dosažené výsledky

- Pipeline od RSS feedu po publikaci
- 5 nezávislých workflow v n8n
- AI scoring + generování v češtině
- Automatická deduplikace
- WordPress publikace s excerpty

Plánovaná rozšíření

- Řídící tabulka (multi-topic)
- Overview dashboard
- Práce s multimédií (featured image, obrázky v těle)
- GA4 + GSC napojení
- Feedback loop (učící se systém)
- Příprava publikace na Ghost CMS
- Vektorová DB — sémantická deduplikace