



# MariaDB Vector: Making AI transparent on your own data

*Use case: Wikipedia editing with RAG*

*foss-north, Göteborg, Sweden  
Fri 14 Apr 2025*

**Robert Silén**

[robert.silen@mariadb.org](mailto:robert.silen@mariadb.org)

Community Advocacy, MariaDB Foundation

# Why transparency on AI on your own data?

- When accuracy, sources and accountability matter for the user
- When you need control and insight to intermediate steps

How: Use an open source database for your AI apps:

- Store all data in the same open source standard RDBMS: data, sources, vectors and intermediate results
- Single queries can combine vector & standard data
- Developers know the standard RDBMSes
- Numerous tools work with the standard RDBMSes
- Lack of vendor lock in
- Lower total cost

# Agenda

1. Use case: RAG for editing Wikipedia
2. What exactly is MariaDB Server?
3. What AI functionality does MariaDB have?
4. What are the main use cases for MariaDB Vector?
5. Steps in creating RAG with MariaDB Vector
6. So where is the advantage of using MariaDB?
7. Technical details (another TOC; on your own time)

# Projekt Fredrika

- Non-profit founded 2017 by Kaj Arnö
  - Improve Swedish Finland on open data
  - Wikipedia, Wikidata, Wikimedia Commons
- 
- Fredrika Runeberg, wife of national poet Johan Ludvig Runeberg.  
Journalist, and first women author of Finland



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB						
1			TOTAL VIEWS BY LANGUAGE (30d/1y):																															
2	Q-kod	Artikel	Typ	Prio	Commons	Svenska	ang_coun	sv_30d	sv_1y	fi_30d	fi_1y	en_30d	en_1y	de_30d	de_1y	fr_30d	fr_1y	ca_30d	ca_1y	cs_30d	cs_1y	es_30d	es_1y	et_30d	et_1y	he_30d	he_1y	hu_30d						
3	Q2815380	Kaj_(humorgrupp)	KAJ	1		<a href="https://sv">https://sv</a>		22	184464	195549	155452	160394	99124	99124	9179	9179	2127	2127	103	103	535	535	2246	2246	582	582	774	774	355					
4	Q1315611	Bara_bada_bastu	KAJ	1		<a href="https://sv">https://sv</a>		15	38905	38905	52303	52303	76262	76262	4999	4999	652	652									621	621	277					
5	Q9624204	Axel_Åhman	KAJ	1	Category:	<a href="https://sv">https://sv</a>		3	80208	82979	23891	23891																						
6	Q1315613	Kevin_Holmström	KAJ	1		<a href="https://sv">https://sv</a>		3	72369	72369	25093	25093																						
7	Q1315530	Jakob_Norrgård	KAJ	1	Category:	<a href="https://sv">https://sv</a>		3	80341	80341	26880	26880																						
8	Q1461092	Finlandssvenska	Språk	1		<a href="https://sv">https://sv</a>		22	5977	24337	0	9108	10100	54574	809	5620			21	104			0	5047										
9	Q815342	Nessun_dorma	Musik	1		<a href="https://sv">https://sv</a>		37	2757	11121	0	6592	27664	459948	4076	79332	1761	34119	0	2991	419	7808	4465	70346		121	1480							
10	Q7654934	Österbottenska	Språk	2		<a href="https://sv">https://sv</a>		5	13134	16382	7984	9146																						
11	Q2156306	Vörå	Ort	2		<a href="https://sv">https://sv</a>		25	38566	43268	74604	86316	23280	26725	1900	2159	285	350	0	0	111	111	227	313	150	150								
12	Q57036	Bastu	Kultur	2		<a href="https://sv">https://sv</a>		72	4995	13998	4363	28150																						
13	Q726673	Finlandssvenskar	Nation	2		<a href="https://sv">https://sv</a>		33	11489	34877	5450	36436	1539																339	639				
14	Q3188515	Professionella_pjasalappar	KAJ	2		<a href="https://sv">https://sv</a>		1	993	1157																				786	80			
15	Q5702	Österbotten_(nutida_landskap)	Ort	2		<a href="https://sv">https://sv</a>		60	0	32073	7186	47272	1172																		281			
16	Q1332333	Taco_hej_(me_Gusta)	KAJ	2		<a href="https://sv">https://sv</a>		1	1099	1099																								
17	Q1334446	Pa_to_ta_na_kako?	KAJ	2		<a href="https://sv">https://sv</a>		1	43	43																								
18	Q9027	Svenska	Språk	3		<a href="https://sv">https://sv</a>		201	0	76456	2542	26566	3317																			755	491	
19	Q33	Finland	Ort	3		<a href="https://sv">https://sv</a>		321	0	118268	46408	515465	21872																			093	5054	
20	Q643535	Österbotten_(historiskt_landskap)	Ort	3		<a href="https://sv">https://sv</a>		23	1384	6014	613	6500	209																				14	
21	Q1068564	Svenska_Österbotten	Ort	3		<a href="https://sv">https://sv</a>		1	4461	8462																								
22	Q1232329	Språk_i_Finland	Språk	3		<a href="https://sv">https://sv</a>		12	0	5764	1031	8689	504																					
23	Q3189004	Kom_ti_byin	KAJ	3		<a href="https://sv">https://sv</a>		1	975	1234																								
24	Q3751292	Svenskifinland	Ort	3		<a href="https://sv">https://sv</a>		6	1901	7457	165	1146	41																					
25	Q125080	Vasa	Ort	4		<a href="https://sv">https://sv</a>		80	0	44156	0	86652	749																				409	134
26	Q1184211	Eurovision_Song_Contest_2025	Musik	4		<a href="https://sv">https://sv</a>		48	28205	60381	39018	101964	78994																				195	7036
27	Q235269	Pernilla_Karlisson	Musik	4		<a href="https://sv">https://sv</a>		27	1698	2184	2780	5485	245																					11
28	Q4261412	Tävlingen_för_ny_musik	Musik	4		<a href="https://sv">https://sv</a>		13	681	2151	3800	45585	253																					34
29	Q945779	När_jag_blundar_(Pernilla_Karlisson-låt)	Musik	4		<a href="https://sv">https://sv</a>		15	0	1529	1379	2314	54																					7
30	Q3351781	Yle_X3M	Kultur	4		<a href="https://sv">https://sv</a>		7	0	1708	507	3284	35																					
31	Q3431686	Wasa_Teater	Kultur	4		<a href="https://sv">https://sv</a>		7	980	1613	415	937	26																					
32	Q1072782	Österbottens_förbund	Kultur	4		<a href="https://sv">https://sv</a>		2	1295	1938	0	311																						
33	Q427597	Arja_Saijonmaa	Musik	4		<a href="https://sv">https://sv</a>		12	0	147049	17525	63337	777																					



## KAJ references

## References [edit]

- <sup>1</sup> <sup>^</sup> [KAJ med i Melodifestivalen – utlovar dialekt och sauna-pop](#) [KAJ in Melodifestivalen – promises dialect and sauna pop] (in Swedish). *Sveriges Radio*. 27 November 2024. Retrieved 24 February 2025.
- <sup>2</sup> <sup>^</sup> ["KAJ" \*ö\*. Nästan unplugged. Series 3. Episode 2 \(in Swedish\). Alexander Theatre: Yle. 21 September 2018. Yle Fem. Retrieved 23 March 2025.](#)
- <sup>3</sup> <sup>^</sup> [Lindberg, Louise \(8 July 2015\). "KAJ + K-pop=sant" \*ö\* \[KAJ + K-pop=true\] \(in Swedish\). Yle. Retrieved 23 March 2025.](#)
- <sup>4</sup> <sup>^</sup> [Aldén, Nicke; Westerlund, Märta \(4 August 2016\). "Så här gjordes KAJs nya musikvideo - i en tagning" \*ö\* \[This is how KAJ's new music video was made - in one take\] \(in Swedish\). Yle. Retrieved 23 March 2025.](#)
- <sup>5</sup> <sup>^</sup> [Wilow, Felix \(13 March 2025\). "Marianne Eklöf Törnbloms hyllning till KAJ-Jakob" \*ö\* \[Marianne Eklöf Törnblom's tribute to KAJ-Jakob\] \(in Swedish\). \*Aftonbladet\*. Retrieved 23 March 2025. \[Jakob Norrgård's performance of "Nessun dorma" with his own text about driving in Finland has become a success on social media with over 600,000 views on Youtube.\]](#)
- <sup>6</sup> <sup>^</sup> ["KAJ: Nissan bromsa \(Nessun dorma\), storslagen aria på Alexandersteatern" \*ö\* \[KAJ: The Nissan braked \(Nessun dorma\), grand aria at the Alexander Theatre\] \(in Swedish\). Yle. 21 September 2019. Retrieved 23 March 2025.](#)
- <sup>7</sup> <sup>^</sup> [Isojoki, Jukka \(13 June 2019\). "Vörjeans är världens bästa rockabillyband från Vörå" \*ö\* \[Vörjeans is the world's best rockabilly band from Vörå\]. \*Yle\* \(in Swedish\). Retrieved 13 April 2025.](#)
- <sup>8</sup> <sup>^</sup> ["Sverige har fått KAJ-feber – Spotifysuccén efter Mellodebuten" \*ö\* \[Sweden has caught KAJ fever – the Spotify hit after the Mellodebut\]. \*Aftonbladet\* \(in Swedish\). 24 February 2025. Retrieved 24 February 2025.](#)
- <sup>9</sup> <sup>^</sup> ["KAJ has won 'Melodifestivalen' and will represent Sweden in Basel" \*ö\*. Eurovision Song Contest. 8 March 2025. Retrieved 8 March 2025.](#)
- <sup>10</sup> <sup>^</sup> ["Axel Åhman om barndomen och fotbollen" \*ö\* \[Axel Åhman on childhood and football\] \(Podcast\). Yle. Retrieved 27 February 2025.](#)
- <sup>11</sup> <sup>^</sup> ["Hur KAJ blev KAJ" \*ö\* \[How KAJ became KAJ\] \(in Swedish\). Yle. Retrieved 27 February 2025.](#)
- <sup>12</sup> <sup>^</sup> ["Kevin Holmström är Vegas vinterpratare" \*ö\* \[Kevin Holmström is Vega's Winter Speaker\] \(in Swedish\). Yle. 20 December 2023. Retrieved 27 February 2025.](#)
- <sup>13</sup> <sup>^</sup> ["Recension: Botnia Paradise är som en läcker godispåse från taxfreebutiken, och ingen förälder stoppar dig och säger att "nu räcker det" " \*ö\* \[Review: Botnia Paradise is like a delicious bag of sweets from the](#)
- <sup>29</sup> <sup>^</sup> [Ninos, Fotios \(11 March 2025\). "Sweden: "Bara Bada Bastu" hits No. 1 on Spotify's Daily Global Viral Chart!" \*ö\*. \*Eurovision Fun\*. Archived \*ö\* from the original on 11 March 2025. Retrieved 11 March 2025.](#)
- <sup>30</sup> <sup>^</sup> [Yli-Ojanperä, Elina \(16 March 2012\). "När jag blundar – Pernilla Karlsson laulaa äidistään veljensä sanoin ja sävelin" \*ö\*. \*Yle Luovat sisällöt ja media\* \(in Finnish\). Retrieved 8 March 2025.](#)
- <sup>31</sup> <sup>^</sup> [Karttunen, Camilla \(11 March 2025\). "KAJ'n kappale rikkoo Euroviisujen sääntöjä – Tässä voi olla ratkaisu" \*ö\*. \*Iltalehti\* \(in Finnish\). Retrieved 11 March 2025.](#)
- <sup>32</sup> <sup>^</sup> [\*a b c d\* Lüttge, Peter; Åhman, Erik; Fant, Izabella \(2 March 2025\). "I Vörå finns i dag bara stolta och lyckliga människor – följ med på en resa genom KAJ-land" \*ö\* \(in Swedish\). Yle. Retrieved 11 March 2025.](#)
- <sup>33</sup> <sup>^</sup> [\*a b c d e f g h i\* Boo, Sofia \(9 March 2025\). "Det här är Kaj – vinnarna av Melodifestivalen 2025" \*ö\*. \*Göteborgs-Posten\* \(in Swedish\). Retrieved 11 March 2025.](#)
- <sup>34</sup> <sup>^</sup> ["Kevin Holmström: "Ingenting är definitivt för mig" \*ö\*. \*Kyrkpressen\* \(in Swedish\). 24 January 2024. Retrieved 11 March 2025.](#)
- <sup>35</sup> <sup>^</sup> [Touur, Ville \(9 March 2025\). "Nämä tunnetut suomalaiset ovat KAJ:n tähden serkkuja" \*ö\*. \*Ilta-Sanomat\* \(in Finnish\). Retrieved 11 March 2025.](#)
- <sup>36</sup> <sup>^</sup> [\*a b\* Popenoe Thor, Clara \(11 March 2025\). "A:et i Kaj kan mer än bara bada bastu" \*ö\*. \*Dagens Nyheter\* \(in Swedish\). Retrieved 11 March 2025.](#)
- <sup>37</sup> <sup>^</sup> ["Vinnarna i SFV:s Arvid Mörne-tävling 2017" \*ö\* \(in Swedish\). \*Svenska folkskolans vänner\*. 8 May 2017. Retrieved 11 March 2025.](#)
- <sup>38</sup> <sup>^</sup> [\*a b\* Koivuranta, Riitta \(10 March 2025\). "Tämä kaikki KAJ:n jäsenistä tiedetään" \*ö\*. \*Helsingin Sanomat\* \(in Finnish\). Retrieved 11 March 2025.](#)
- <sup>39</sup> <sup>^</sup> [Lempinen, Jenna \(10 March 2025\). "Tämä kaikki KAJ-kolmikosta tiedetään – pitkiä parisuhteita ja yllättävät taustat" \*ö\*. \*Ilta-Sanomat\* \(in Finnish\). Retrieved 11 March 2025.](#)
- <sup>40</sup> <sup>^</sup> [Andrée, Benjamin \(8 March 2025\). "KAJ:s okända kärleksliv: "Mönster i gruppen" " \*ö\*. \*Expressen\* \(in Swedish\). Retrieved 11 March 2025.](#)
- <sup>41</sup> <sup>^</sup> [\*a b c\* Holmberg, Bianca \(16 June 2016\). "Humorn är grunden – dialekten skatten för KAJ" \*ö\* \[Humour is the foundation – dialect the treasure for KAJ\] \(in Swedish\). Språkbruk. Retrieved 13 April 2025.](#)
- <sup>42</sup> <sup>^</sup> ["Jakob charmade Kevin och Axel genom att hjula och slå volter - sedan föddes KAJ" \*ö\* \[Jakob won over Kevin and Axel with his cartwheels and flips – and thus, KAJ came into being\] \(in Swedish\). Yle. 10 November 2017. Retrieved 13 April 2025.](#)



# Uppslagsverket Nagu

- Nagu on Wikipedia - Fredrika's first project
- Archipelago, South of Åbo
- Improved hundreds of articles with text and images, and over a thousand islands
- Organized "skrivstugor" with local experts
- Results in Wikipedia were good enough for a physical printed encyclopedia



# From themes and experts, to sources

- Reliant on experts, and even more so, on sources
- Fredrika's Wikipedian in Residence
  - [150 000+](#) edits on Wikidata
  - [10 000+](#) edits on 4500+ Wikipedia articles
  - [100+](#) articles created on Swedish Wikipedia
- How to scale efforts?

The screenshot shows the Project Fredrika website with a grid of news articles. The top navigation bar includes the Project Fredrika logo and a list of regions: ÅLAND, ÅNGÅ, INGÅ, IRORETT, FÖRFATTARE, WIKIPEDIA, MOTEN, TILBÄRBRUCK, FINLAND, UKRAINA, HELSINGFORS, BASEBORG, BIRGER CARLSTEDT, ÖVRETT. The main content area is a grid of 12 article cards. Each card has a title, a date, a small image, and a short text snippet. The articles cover various topics such as 'En tillbakablick på 2024', 'Fredrika matar AI-draken', 'AI Saarna: Improving Wikipedia with Gen AI', 'Projekt Fredrika mot ett AI-fyllt år 2024', 'Sommarpratet 2023', 'Intervju med Jan Sten: att redigera Wikipedia var ju inte så svårt', 'Roberts mest lästa 2022', 'SLS material när ny publik via Wikipedia', 'Ny hemsida med Hugo', 'Lyckade skrivstugor: Våra lärdomar', 'Bidrag till Ukraina på Wikipedia', 'Våra mål', 'Finlandsvenska tekniker', 'Ukrainas historia på Wikipedia', 'Nato-intresset större i Finland och Sverige än inom Nato-länderna själva', 'Källor på Wikipedia', 'ASA-uppdrag', and 'Finlands skärgård på Wikipedia'. The bottom left corner of the screenshot shows the MariaDB Foundation logo.





# Vision: RAG for Wikipedia editing based on sources

- A human-AI co-editing system where AI helps identify gaps, biases, and outdated information in articles
- Suggests improvements grounded in verifiable, high-quality **sources**
- Editors remain in control – AI becomes a transparent, explainable tool, not an autonomous agent with vetting queues for AI-suggested changes
- Respect existing Wikipedia guidelines as the main priority, in particular NPOV, encyclopaedic language style and high-quality source references

# RAG for Wikipedia editing

1. Start from identifying **great sources** (high quality books, articles).
2. Prepare by identifying which people, places and other Wikidata Q codes the source material is about

By using **NER** (*Named Entity Recognition*) and with **vectorization**: find closest matching chunks of source and Wikipedia articles by creating vectorized embeddings of them.

3. Generate Wikipedia improvement suggestions with **RAG** (*Retrieval Augmented Generation*): **prompt** an LLM with the source and wikipedia article, with a task and required output.
4. Review suggestions, and add to Wikipedia with the **source reference**.



# What was the result?

- We found 278 unique people och 265 places in the book
- We generated Wikipedia improvements suggestions on most mentioned
- The author of the book evaluated and
  - approved  $\frac{1}{3}$  as such to be added to Wikipedia,
  - approved  $\frac{1}{3}$  after manual modifications, and
  - dismissed  $\frac{1}{3}$
- Challenges
  - Matching content to Wikipedia:  
is “Erik” king Erik XIV or Edelfelt’s son?  
is “Alexandra” queen Alexandra of Denmark, or Albert’s mother?
  - Filtering out bad suggestion:  
give AI the option to say “sources are not good enough”



# How was AI used in this use case?

- Wrote python scripts using AI Cursor and Claude
- Chunkification of ebook to chapters, pages, paragraphs
- NER: SpaCy large model in Swedish combined with “entitiy fishing”
  - <https://spacy.io/models/sv>
  - <https://github.com/kermitt2/entity-fishing>
- Generated wikipedia suggestions with
  - Anthropic Claude - best objective output!
  - OpenAI, ChatGPT
  - Mistral
  - Google Gemini



# Insights

- **Chunkifications matters**
  - One vector per document/article, chapter, page, paragraph or sentence
  - Most relevant chunk  $\neq$  chunks fed to LLM
- **Vectorize whole Wikipedia?**
  - All articles in all languages is a 76 GB zip file
  - MariaDB can handle the vectors, vector embedding a subset is cheaper in time and money
  - Pre Vectorized datasets can be available
- [https://sv.wikipedia.org/wiki/Wikipedia:Projekt\\_Fredrika/SLS-AI-pilot](https://sv.wikipedia.org/wiki/Wikipedia:Projekt_Fredrika/SLS-AI-pilot)





# Future of RAG

- New, improved LLMs with less hallucination and better focus
- Do It Yourself (DIY) vs AI frameworks
- AI RAG Frameworks
  - Better prompts
  - Better chunkification strategy
  - AI-agent workflows
  - Graph database approach to extract claims
  - Multi-hop looping - refine user query before vector search
- Every source is unique and RAG will need to be adopted for optimal results

# Thank you!

Contact details:

[robert.silen@mariadb.org](mailto:robert.silen@mariadb.org)

About:

<https://mariadb.org/projects/mariadb-vector/>

[https://sv.wikipedia.org/wiki/Wikipedia:Projekt\\_Fredrika/SLS-AI-pilot](https://sv.wikipedia.org/wiki/Wikipedia:Projekt_Fredrika/SLS-AI-pilot)