

Docker container · Azure Container Apps

Frontend (single-page HTML / JavaScript)

Tabs: Concordancer · Frequency · Markers · Text analysis
Register comparison · Regex · LLM · Chart.js (CDN)

REST API (16 endpoints, JSON)

Backend · FastAPI (Python, async / aiosqlite)

Query-time text processing:

tokenization · frequency · n-grams
MI collocations · log-ratio keyness

SQL · FTS5 full-text queries

Optional LLM layer

BYO key · Gemini / OpenRouter

SQLite databases · FTS5 (unicode61)

corpus 472.9M words

literature 18.5M words

dialectal 5.9M words

twitter 1.9M words

clarin_crete 0.12M words

Pre-computed

marker_
frequencies

93 markers
× registers

Databases mounted from Azure File Share (decoupled from image)