



DATA ANALYTICS

# CLAVIN

Open source software for document geolocation parsing, tagging, and resolving.

CLAVIN (Cartographic Location and Vicinity Indexer) is an open source software package for geotagging and geoparsing. It automatically extracts location names from structured and unstructured text and resolves them against a gazetteer to produce data-rich geographic entities. CLAVIN is fast, accurate, and scalable to accommodate big data in the cloud. CLAVIN combines various open source tools with natural language processing (NLP) techniques to extract and resolve geospatial entities while reconciling misspellings, alternate names, and ambiguous references. By enriching documents with structured geodata, CLAVIN enables low-cost, advanced geospatial analytics on massive volumes of text.

---

## ADDRESSING THE NEED

As the demand for geospatial analytics continues to grow, most human knowledge remains trapped in unstructured text. Competing solutions require costly enterprise licenses and often charge by the document, quickly becoming cost-prohibitive. As big data grows bigger, you can't afford to pay by the byte, document, or CPU.



Geotag one million documents per hour on a small Hadoop cluster of commodity servers.

# Retrieve rich geospatial data from structured and unstructured text.

---

## THE CLAVIN ADVANTAGE:

### Intelligent

- Use NLP, fuzzy matching, and context-based heuristics.
- Recognize alternative names as referring to the same geographic entity, utilizing fuzzy search to reconcile incorrect spellings.

### Accurate

- 75% accuracy for geospatial entity resolution.
- Precisely resolve to the geographic entity intended by the author based on document context.

### Versatile

- Advanced geospatial analytics and hierarchical geospatial search integrated into multiple next-generation analytic platforms.
- Render results in simple API in virtually any visualization tool.

### Disruptive

- Replace costly proprietary tools, providing substantial cost savings while increasing flexibility.
- Process 1 million documents in under an hour on a 9-node Hadoop cluster.
- Scale to billions of records.

### Customizable

- Flexible licenses and advanced options for custom gazetteers, integration, and support.



Automatically resolve 100 locations per second, per CPU at 75% accuracy.