

Overview of Whoosh and Seaborn

Nick Collins

Data Rave 2014-04-02

Whoosh

- **Pure Python** search engine library
- Written by **Matt Chaput** for use in Houdini (3D animation software)
- **Just a library**, not client-server (unlike Elasticsearch, Solr, Lucene).

Whoosh – key components

- **Schema and Index**
 - Index
 - Schema specifies the fields of the index
- **IndexWriter** – used to add documents
- **Searcher**

Whoosh - querying

- Default parser is the **QueryParser** object
- **Search** object takes a parsed query and returns a **Result** containing **Hits**.

Whoosh - query language

- Phrases can specify word distance
 - “data meetup”~5 matches “data rave NYC's April meetup”
- Boolean terms: “data AND rave”
- Fields: “title:data”
- Wildcards: * and ?
- Ranges: “date:[20050101 TO 20090715]”
- Boosting: “date^0.5 rave^2”

Seaborn

- A graphics library based on matplotlib adding:
 - Built in themes and color palette tools
 - Integrated plotting and analysis, fits models and displays uncertainty in the plot
- Written by Michael Waskom, a psychology grad student at Stanford.

Links

- <https://bitbucket.org/mchaput/whoosh/wiki/Home>
- <http://www.stanford.edu/~mwaskom/software/seaborn/index.html>