Project re-Iseach Summary Sheet v.3

Contact: Edward C. Zimmermann <edz@nonmonotonic.net>
LinkedIn: https://www.linkedin.com/in/edwardzimmermann/

Project re-isearch: a novel multimodal search and retrieval engine using mathematical models and algorithms different from the all-too-common inverted index. The design allows it to have effectively no limits on the frequency of words, term length, number of fields or complexity of structured data and support even overlap—where fields or structures cross other's boundaries (common examples are quotes, line/sentences, biblical verse, annotations). Its model enables a completely flexible unit of retrieval and modes of search.

- ✔ Low-code ETL / "Any-to-Any" NoSQL datastore architecture
- ✔ Handles a wide range of native (and via filters) document formats including “live” data.
- ✔ Phrase, case, proximity, wildcard, parametric, range, phonetic, fuzzy, thesauri, polymorphism datatypes (including numeric, dates, geospatial, ranges etc.) and object capabilities.
- ✔ Set based with an exhaustive collection of (binary and unary) set operations.
- ✔ A number of different query languages including “smart” plain language expressions.
- ✔ Fully Customizable and extendable.
- ✔ Plugin architecture that allows for binary distributed (proprietary) 3rd party extensions
- ✔ Useful for Analytics, Recommendation / Autosuggestion and a host of other applications
- ✔ Support for Peer-to-Peer and Federated architectures.
- ✔ Flexible scripting language interfaces (including Python)
- ✔ Tiny, efficient. Can run on low powered systems (even embedded) with a minimum of RAM
- ✔ Freely available under a permissive software license.

Core engine development language: reduced subset of highly portable C++
Plugin extensions development language: C++
Application development language: Among others C++, C, Java, PHP, **Python**, R, Tcl/Tk
Licence: Apache 2.0

A few possible paradigm changing uses:

- Multi-media/video: using a combination of speech to text and image captioning pre-procesing.
- distributed internet search on IPFS
- Centroid federated search

Source: https://github.com/re-Isearch/re-Isearch

Comparisons to other engines (Especially Lucene):

https://github.com/re-Isearch/re-Isearch/blob/master/docs/re-Isearch-vs-Others.pdf

Handbook:

https://github.com/re-Isearch/re-Isearch/blob/master/docs/re-Isearch-Handbook.pdf

Design Whitepaper:

https://github.com/re-Isearch/re-Isearch/blob/master/docs/re-Isearch-Design.pdf

FOSDEM’22 Talk:

https://fosdem.org/2022/schedule/event/lt_re_lsearch/ (~14 min)

Project re-Iseach Summary Sheet v.3
Contact: Edward C. Zimmermann <edz@nonmonotonic.net>
LinkedIn: https://www.linkedin.com/in/edwardzimmermann/

re-Isearch

References:

European Union N(ext) G(eneration) I(nternet): https://www.ngi.eu/funded_solution/re-isearch/
Nlnet Foundation: https://nlnet.nl/project/Re-iSearch/
Visual Search: https://isea-archives.siggraph.org/art-events/metahaven-exodus-cross-search/

Comparisons

	re-Isearch	Typesense	Algolia	ElasticSearch	Meilisearch	intraFind
Open Source?	Yes	Yes	No	Source-available	Yes	No
License	Apache 2.0	GPL 2	Commerical	SSPL	MIT	Commercial
First Commit	1992,2020	2015	2012	2010	2018	2010
Built Using	C++	C++	C++	Java	Rust	Java
Core Search Algorithm	Own	Own	Own	Lucene	Own	Lucene
Primary Index Location	Disk, exploits virtual memory system	RAM	RAM	Disk, with RAM cache	Disk with Memory Mapped files	Disk, with RAM cache

re-Isearch	MarkLogic	Elasticsearch	Apache Solr
NoSQL search engine	Operational and transactional Enterprise NoSQL database	A distributed, RESTful modern search and analytics engine	A widely used distributed, scalable search engine
NativeXML DBMS, RDF Store, search engine	NativeXML DBMS, RDF Store, search engine	Search engine	Search engine
Object DBMS including Spatial		Document store Spatial DBMS	Spatial DBMS
1994-2011, reborn 2021	Since 2001	Since 2010	Since 2006
C++	C++	Java	Java
Free	Commerical	Partially Free	Free
Open Source	Proprietary	Open Source	Open Source
XML support	XML support	JSON Only	XML support
Foreign keys, Join	No foreign keys	No foreign keys	No foreign keys
Schema-Free	Schema-Free	Schema-Free	Schema
Multi-language API, Z39.50, SRU/W. CQL, IB Query Language,...	Multi-language API, Xquery, SPARQL,	Java API RESTful HTTP/JSON API	Java API RESTful HTTP/JSON API
Search during index	Search during index	No	No
Own algorithms	Own algorithms	Based on Lucene	Based on Lucene

Project re-Iseach Summary Sheet v.3
Contact: Edward C. Zimmermann <edz@nonmonotonic.net>
LinkedIn: https://www.linkedin.com/in/edwardzimmermann/

Market Overview

There are many of the companies in this space and market caps are relatively high. Intrafind and Elastic are based on Apache Lucene.

Figure 1: Magic Quadrant for Insight Engines

Source: Gartner (March 2021)

Project re-Iseach Summary Sheet v.3
Contact: Edward C. Zimmermann <edz@nonmonotonic.net>
LinkedIn: https://www.linkedin.com/in/edwardzimmermann/

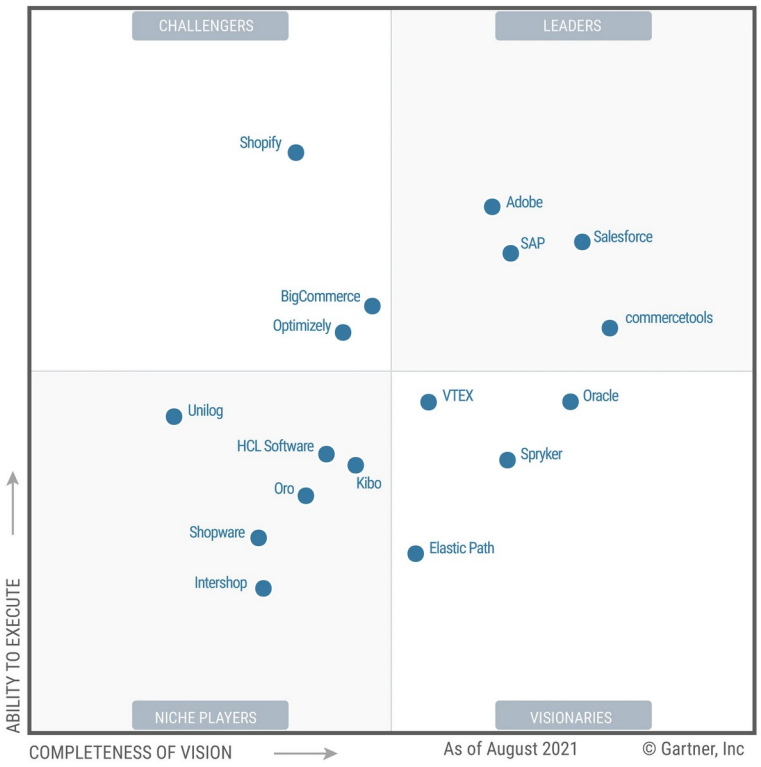
re-Isearch

Commercial market: ecommerce product search.

The engine has a number of unique features that set its possibilites apart from the standard solutions.

Elastic Path's Java ecommerce platform is based on open source technologies such as Spring Framework, Apache OpenJPA, Eclipse RCP, Apache Solr, Apache Velocity, Groovy, Direct Web Remoting, jQuery and more. So basically … Lucene….

Figure 1: Magic Quadrant for Digital Commerce



Source: Gartner (August 2021)