

**Edward C. Zimmermann <[edz@nonmonotonic.net](mailto:edz@nonmonotonic.net)>**  
**<http://www.nonmonotonic.net/re-isearch>**

# Project re-Search

The 27 year old new kid on the search block



# History

- Isearch was a legendary open-source text retrieval software first developed in 1994 as part of the Isite Z39.50 information framework with support from NSF.
- Development was divided between CNDIR/MCNC and Bsn (Germany).
- In 1998 BSn launched a proprietary fork using new algorithms.
- It was deployed in 100s of high profile sites....
- In 2011 Bsn/NONMONOTONIC's proprietary fork ceased development and it moved to the attic..

***Despite lack of support some servers are still running!???***

***Don't break a working system.....***

# You mentioned sites??? Curious!

The U.S. Patent and Trademark Office (USPTO) patent search, the Federal Geographic Data Clearinghouse (FGDC), the NASA Global Change Master Directory, the NASA EOS Guide System, the NASA Catalog Interoperability Project, the astronomical pre-print service based at the Space Telescope Science Institute, The PCT Electronic Gazette at the World Intellectual Property Organization (WIPO), the SAGE Project of the Special Collections Department at Emory University, Eco Companion Australasia (an environmental geospatial resources catalog), the Open Directory Project, genomic search for the Australian National Genomic Information Service's (ANGIS) human genome project (and its eBiotechnology workbench split-off); the D-A-S-H search portal against racism, antisemitism and exclusion (funded within the framework of the action program "Youth for tolerance and democracy - against right-wing extremism, xenophobia and anti-Semitism", the YOUTH program of the European Community and with additional support from the German Federal Agency for Civic Education); the e-government search (Yeehaw) of the U.S. State of Utah to agronomic cooperation across the Mediterranean region (supported by the EU's DG). Integrated into a number of CMS platforms it powered search for a number of high volume web sites. It was also used as a database accelerator by a number of eCommerce shops.

**Development had been terminated.. Loads of people asked if I could open source IB.. and we were constantly surprized at ApacheCon just how primitive the offerings were..**

*And thanks to the kind support of NInet/NGI0...*

*Reborn in 2020 in the middle of the global Covid19 pandemic as Project re-Issearch. Like the original, it is not just about textual words but pushes the envelope.*

# Motivation. What is different?

Mainstream search engines are about finding any information:

*"a list of all documents containing a specific word or phrase".*

So search engines paradoxically return both too much information (i.e. long lists of links) and too little information (i.e. links to content, not content itself).

The re-Isearch engine is, by contrast, about exploiting document structure, both implicit (XML and other markup) and explicit (visual groupings such as paragraph), to zero in on relevant sections of documents, not just links to documents.

We call it: **Smart queries and run-time (dynamic) unit of retrieval.**

# What do you mean?

In "traditional" search engine models there is a standard unit of the record. Its the unit of index (the page, PDF, Word document) and retrieval. By contrast we have a user defined "search time" unit of retrieval: the structure of documents can be exploited to identify which document elements (such as the appropriate chapter or page) to retrieve. Retrieval granularity may be on the level of sub-structures of a given document or page such as line, paragraph but may also be as part of a larger collection. Since we know the location of hits (matches) within a record we can transverse its structure (which can be viewed as a graph) to find other relevant bits or retrieve relevant elements. These can even be virtual.

Pretty cool, eh!?

# What do you mean by virtual?

Instead of just searching for results one can search for clusters of elements (we call them “semantic spheres”) that are relevant.

Words derive their meaning in context. In a social network these are, for example, what people call “bubbles”.

One can dig deeper into these to find new insights and discover different views and aspects (and not just some monolithic “Big Brother” mediated message).

We call it multi-modal *re*-search!



# *Exodus:Cross Search ISEA 08 Singapore*

*Artist: Metahaven (NL)  
Collaborators*

*Tsila Hassine*

*Maurits de Bruijn*

*Edward Zimmerman*

*Norbert Poëllmann*

*Florian Schneider - KEING.ORG*

*Zhou Xiangdong - Fudan University*

# You said “Beyond Text” ?

Like the original, it is not just about textual words but the design contains a large number of objects: numerical, range, geospatial etc.

These objects don't even have to be part of any document but may be available via interface glue into other systems via ODBC, CORBA or object embedding. This allows indexing content to be stored in and searched from other systems. This is useful in many dynamic applications in commerce and trading...

OK.. I'm impressed.. But unique?

# What other uses for objects ?

We support something called "**Dynamic Presentation**".

- When things get indexed (ingested) the structure gets stored with pointers to content. So we can not just search but also reconstruct formats on-the-fly at time of presentation without re-parsing. An e-mail, for example, can be viewed as not just e-mail but als XML, as HTML as .. but there is more..
- Content can come (pointers) from those other object services, remote databases etc. merging into a presentation. It can also (rights based) provide different views (hiding or redacting content).

# Any bullet points?

- Low-code ETL / "Any-to-Any" architecture (Dynamic Presentation)
- Handles a wide range of document formats including "live" data.
- Powerful Search (Structure, Objects, Spatial) / Relevancy Engine
- NoSQL Datastore
- Set based with an exhaustive collection of (binary and unary) set operations.
- Dynamic search-time virtual collections of indexes (Shards)..
- Useful for Analytics, Recommendation / Autosuggestion and ...
- Embeddable in products (comparatively low resource demands)
- Customization.
- Support Peer-to-Peer and Federated architectures (e.g. OASIS SRU/W, ISO 23950/Z.39.50).
- Freely available under a permissive software license (Apache 2.0).

That is really a lot.. Any more features we may have of dreamt of?

Of course. We've barely touched on them (Can go on for hours)....

***Make that weeks..***

*Visit: <http://www.nonmonotonic.net/re-isearch>  
To learn more.*

*Software, documentation and handbooks are freely available on  
Github:*

