



FLAX turns search upside down for Infomedia

Infomedia (www.infomedia.dk) is a leading Danish provider of media intelligence, i.e. media search, media monitoring and media analysis. In 2013 Flax started a project with Infomedia to replace and upgrade their existing media monitoring and media archive systems, based on the Verity and Autonomy IDOL engines, to a modern and scalable open source system. The existing architecture was at full capacity with problems including slow searches and a complex workflow.

Around 17,000 stored queries, representing the interests of Infomedia's clients, had been captured in the legacy Verity VQL syntax and would require translation to a new internal format that would not tie Infomedia to any particular vendor or technology. The new system would also have to scale easily to handle up to 85 million news items in a searchable archive and to

monitor at least 20 new articles every second, with an eventual target of 100 articles a second.

Flax designed a new monitoring service built on the Luwak library, an open source stored query engine based on Apache Lucene and already in use at several other Flax client sites. Luwak allows for many tens of thousands of stored queries to be applied to a new item every second and can be used in monitoring, tagging and classification systems, achieving this performance by first identifying candidate queries by searching all the stored queries with each new document - turning the standard search paradigm upside down. The system scales easily and linearly with the addition of more servers. Flax also helped Infomedia design and implement a searchable archive based on the Apache Lucene/Solr search server, and assisted Infomedia with

the translation of their stored queries from Verity VQL format to a new, internal format, IQL.

The new monitoring system runs on only two standard servers and easily handles the daily update of news items, indicating exactly why and where particular stored queries have matched. The archive search system runs on a further eight servers including an automatic failover capability. Extensive testing of the new architecture was carried out and the monitoring and archive systems entered production in Autumn 2014. Infomedia's new platform allows for easy and economical scalability as query and news volumes increase, prevents vendor lock-in and takes advantage of stable and feature-rich open source software used by thousands of organisations across the world. <https://github.com/flaxsearch/luwak>

BioSolr funded by the BBSRC to improve bioinformatics search

Flax and the European Bioinformatics Institute have recently started a new initiative, BioSolr, funded by the Biotechnology and Biological Sciences Research Council.



The funding will allow Flax staff to work directly with bioinformaticians at the EBI, based at the Wellcome Genome Campus near Cambridge, on the development of new software based on the open source Apache Solr platform. Flax and the EBI will also collaborate on producing workshops, a Solr user group and various publications as well as building links with other groups worldwide interested in improving the search software available in this fast moving and important field of science. The project began in September 2014 with a workshop



day held at the EBI and attended by over 60 people both from the Wellcome Campus and other institutions in the UK and abroad. For further information and to get involved in this important initiative see <https://github.com/flaxsearch/BioSolr>

Flax starts the London Lucene/Solr User Group

The first meeting of a new Lucene/Solr User Group for London is on November 28th at 7pm at Bloomberg's offices in Finsbury Square.

Lucene/Solr committer Shalin Mangar will be taking us on a SolrCloud Deep Dive and Flax will also be presenting on Search Turned Upside Down and on a performance comparison between Solr and Elasticsearch.

The Meetups are free to attend - see <http://www.meetup.com/Apache-Lucene-Solr-London-User-Group> for further details.

About Flax

Flax are independent specialists in open source search software with international clients from sectors including academia, public relations, media, e-commerce and government, including The Financial Times, NLA media access, The University of Cambridge and Reed Specialist Recruitment.

Flax use the power of open source software to drive down costs and provide world beating search performance with no software licence fees.

Flax is an authorized partner of Lucidworks, the commercial company dedicated to Lucene & Solr.