Newsletter

November 2015

www.flax.co.uk

info@flax.co.uk

news

flax

The Open Source Search Specialists

Building an open source search team is hard - let Flax help!

A common complaint from our clients is how difficult it is to find staff with experience of search applications and in particular Apache Lucene/Solr or Elasticsearch. With the explosive growth of open source search over the last few years, there simply aren't enough people on the market with the right skills and experience. We know of some projects that have been looking for staff for over six months!

You might think this is great news for Flax, as we'll always be in demand, but it's not quite that simple. As consultants we're often asked to develop proofs-of-concept or to work as part of a team developing production systems, however our clients will also need to develop search engine skills internally so the project can be supported and continue to evolve. We're also keen to drive development of the core technologies as part of our commitment to the open source community – which we already support by running Meetups, Hackdays and Workshops.

We can help in several ways:

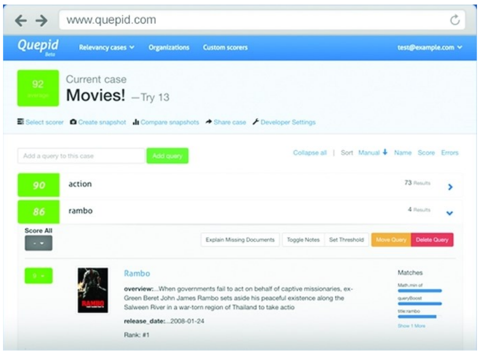
1. Helping to identify internal and external candidates with the right base skills and experience
2. Delivering formal training on Solr and Elasticsearch, basic and advanced topics (either on-site or classroom based)
3. Mentoring focused on a particular project – for example, we might lead a team or individual through some examples in the morning, then they try these techniques on their own data in the afternoon, repeating the process for a week or two – a great way to bootstrap a search project!
4. On-site training for business analysts, managers and others on what is possible with search engines and how to think about querying and analysing your data
5. Providing an introduction to the wider community of search – which books to read, which mailing lists to join and which events to attend

Our experts have years of experience in both software training and open source search and are highly active in the open source community. Do let us know if we can help build your search team.

The real secret to better search…
Quepid, now available from Flax

The real secret to better search is simple. It requires arming developers with the knowledge of those most familiar with the users – marketing, content curators, and domain experts.

Poor interaction between developers and content experts causes search quality to slide backwards. Content experts have little insight into how or why search engines behave. Developers, on the other hand, lack understanding of what search should do. Despite their technical expertise, they rarely have the skills to know what "relevant" means for your application and business. Tuning relevance is therefore a difficult process – fix one problem and it might have a negative affect somewhere else!



overview:...When governments fail to act on behalf of captive missionaries, ex-Green Beret John James Rambo sets aside his peaceful existence along the Salween River in a war-torn region of Thailand to take actio

release_date:...2008-01-24

Rank: #1

Our partners Open Source Connections have built Quepid, a powerful tool for recording and displaying relevance for thousands of queries.

Karen Renshaw, Digital Product Manager – Search Relevancy, for Flax's client RS Components: "Quepid helped us develop best practice approaches across our global search relevancy programme. Using Quepid we are now able to quickly and easily iteratively test and understand the impact of search changes across our entire range of search queries ensuring we make data driven decisions - crucial for organisations where search is critical to the customer experience."

Flax tunes up Elasticsearch for Hadoop for 40x faster indexing

ARACHNYS

Flax recently worked for Arachnys who provide data on a wide range of emerging markets. Their data, gathered by a complex process of web crawling, is stored in HBase and served out of a 10-node Elasticsearch cluster. Indexing all of the 1.2 billion documents would take up to a month! Flax first rewrote how data was mapped from HBase and then identified the single threaded elasticsearch-hadoop plugin as a bottleneck. We replaced this with our own Hadoop OutputFormat implementation using an Elasticsearch TransportClient and BulkProcessor. The BulkProcessor uses a threaded model that sends bulk requests to Elasticsearch using a configurable number of background threads, meaning that there is little or no time spent waiting for responses. Indexing all of the 73 terabytes of data now takes only 17 hours - that's over a gigabyte a second!