flax builds powerful news monitoring solution for Durrants

Durrants Ltd, founded in 1880, is the UK's leading media monitoring company with over 6000 clients from every sector of the economy. The company monitors 10,000 UK sources continously (plus sources from over 100 other countries) and scans over 1.5 million news pages a month.

With such a high volume of incoming content it is imperative that Durrants is at the cutting edge of news monitoring technology. With this in mind they asked text search technology experts Flax to completely replace their existing search engine with an open source solution.

The project had several main aims: to significantly reduce the hardware budget, to reduce the load on Durrants' existing relational database, to improve the quality of results delivered to clients and to provide a stable platform for the development of new features.

The content acquired by Durrants must be monitored against multiple 'agreements' created for their thousands of clients. These agreements can be extremely complex and may run to several pages of logical expressions. Each client may also require Durrants to monitor or exclude various combinations of publications and hierarchical publication types.

Once stories that may be relevant to a client have been found, these are passed to staff for checking and verification. It is obvious that the fewer 'false positives' (stories that do not actually apply to the client's profile) that are found the better, and that 'false negatives' (stories that are relevant but are missed) should never occur.

The new search engine had to match as closely as possible the features of the existing system, so that Durrants' valuable client agreements could be re-used. Durrants were also keen to add new features to improve further the quality of results sent to clients.

Stephen Wicks, CTO of Durrants, said *"We considered several leading commercial vendors of search software, but eventually decided that Flax's open source solution was the most appropriate in terms of performance, flexibility and cost. The Flax team's experience in developing custom search applications was also an important factor, as we knew this project would present unique challenges in terms of scaling and accuracy."*

Flax selected the Xapian search engine and the Python language as key parts of the new monitoring platform. Xapian is a highly scalable open source search library, originally developed in Cambridge, UK as a foundation for a half-billion-page web search engine and featuring Bayesian probabilistic ranking. The Python language allows for rapid development and high performance. Combining these two projects, the Flax team built a client-server architecture that can continue to scale as Durrants' business grows.

The new system supports a detailed query syntax that supersedes and enhances the previous engine's features. Boolean operators, fuzzy matching, word weighting and support for punctuation are all available. The system even allows for errors in scanned content, with matching for known OCR errors. Working closely with the team at Durrants who build client profiles, the Flax team identified cases where certain client requirements produced large numbers of false positives and have reduced this to a minimum: in some cases where the previous engine produced 95% false positives, results from the Flax solution are 95% correct with only 5% false positives.

Further developments have included Web Service APIs to allow the search engine to be tightly integrated into Durrants' workflow and reporting systems. A searchable archive database allows new and existing client agreements to be tested and adjusted to achieve the high performance Durrants' clients demand. Further integration with Durrants' systems is planned.

Stephen Wicks of Durrants said *"The new Flax system runs on virtual machines, and we're currently seeing a reduction in hardware budget to 15% of previous levels. As Durrants continues to grow as a business we now have a predictable and cost-effective way of scaling up the search engine – something that would not have been possible with a closed source solution. We can also easily develop new features to support our internal teams"*.

The Flax team continue to work with Durrants on enhancing and improving the search solution. Charlie Hull of Flax says *"The Durrants project is a great example of how open source search technology can be used to solve problems outside traditional site or intranet search. We've used a search engine as the foundation for a complete media-monitoring solution, and open source software has allowed us to build a highly scalable, accurate and extensible system. We look forward to our continuing relationship with Durrants and to helping them maintain their market-leading position."*

About Flax

Flax is highly active in the information retrieval market with international clients from sectors including academia, public relations, e-commerce, government and private businesses. As a cross-section Flax's clients include The Financial Times, Accenture, the Newspaper Licensing Authority (NLA), The University of Cambridge and Mydeco.

Flax delivers a cutting-edge enterprise search solution, using 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 Lucid Imagination, the commercial company behind Lucene & Solr.

Flax's accolades include an award from the British Computing Society for its powerful, innovative search tools.

For further information please see http://www.flax.co.uk/