



## Lemur Backed by DTI Award to Develop £1.5million Next Generation Search Engine

Lemur is part of a consortium that has been awarded a Department of Trade and Industry grant of £815,000 to build a next generation web search engine.

Early web search engines, such as AltaVista, simply presented a list of documents which contained the words given in the query. Second generation engines, for example Google, added other relevance criteria such as page rank when returning a list of web pages.

Now Lemur is helping to develop the third generation of web search engines. These will include deeper criteria for defining the relevance of the results returned. New factors reveal the intention of web page creators and users. For example, is the page designed to review, or sell? Is the person making the query interested in

researching, or buying?

With over ten years experience of specialising in all aspects of data management, Lemur's software team is active in the development and support of many of today's leading information retrieval products. This placed the company in a strong position to succeed in the DTI-led Technology Programme competition for a £1.5million web search engine project to be developed over two years.

Welcoming the new partnership, Science and Innovation Minister, Malcolm Wicks said: "The UK has a proud history of innovation in science and technology. We believe that we must work with industry to develop the marketable products and services of tomorrow, so that we can maintain our position as a leading global economy.

"That's why we're

supporting this project to develop the next generation of search engines. It provides a great opportunity to harness the UK's world-class expertise and use it to help boost our economy."

**DTI**  
Technology Programme  
**AWARD**  
for  
Third Generation Search

Commenting on this latest award, Lemur's Managing Director, Charlie Hull, said "We were delighted to get through the two highly competitive application rounds to be awarded this prestigious grant. These plans for the next generation of web search engine are very exciting and we are looking forward to working with an interesting group of organisations to achieve them."

### Stemming

Information Retrieval is essentially a matter of finding specific documents in a collection which relate to a user's query. Queries contain one or more search terms (words) which are compared with an index of terms compiled from the collection of documents.

However, words occurring in documents and queries can have morphological variants, for example singulars and plurals, or other grammatical forms, e.g. *fishing*, *fished*, *fisher* and *fish*.

To retrieve documents relevant to search terms and not miss the morphological variants which are likely to have similar semantic interpretations, more accurate search engines incorporate stemming algorithms. These work by reducing the words in the queries and documents to their stem, or root, forms (e.g. *fish* in the example above) so that all relevant documents are found.

As well as increasing relevance, this has the added benefit of reducing the number of distinct terms needed to be stored in the index saving disk storage space and search time.

Lemur's Richard Boulton contributes to the Snowball Project which promotes and develops stemming algorithms, available in 14 languages. For more information see:

<http://snowball.tartarus.org/>

### About Lemur

Lemur is highly active in the information retrieval market with international clients from sectors including academia, public relations, e-commerce, government and private businesses.

Lemur's Flax™ project 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.

Lemur's Bamboo™ software is a collaborative, web-enabled, multimedia database system used by academics and researchers.

Lemur also provides general expertise and development in technologies such as Java, C++, Linux, Windows and XML.