



Mining the deep web for business

► InfoClew – building time-dependent information networks from the deep web

Academic: Dr Bogdan Vrusias

Web searches conducted by businesses often fail to find the information they require because of a lack of sophistication in the mining tools available. Surrey researchers have developed a new technique that is faster, more accurate and cost-effective.

Despite recent efforts by the major search engine companies to introduce clever ways to retrieve meaningful information from the web, searches conducted by businesses are still problematic because data is unstructured and often hidden in the 'deep' web. This IAA project aimed to fill the gap in the market, enabling users to locate, retrieve and visualise meaningful contextualised information about businesses, and present this information concisely and coherently as the company's 'profile'.

The project was co-funded by the IAA, Technology Strategy Board (now Innovate UK)

and local technology company Technotomy. Surrey academics and Technotomy worked with UK Trade & Investment and FDI Marketing Associates to develop and implement a concept for the foreign direct investment (FDI) market.

The system they have developed identifies potential foreign investors in a cost-effective way that is automatic, visual and offers accurate, current and relevant information. The system can search the web to identify candidate companies of a certain profile, for example, those that are likely to expand their operations into a foreign location.

These techniques can easily be applied to other business requirements, leading to numerous potential applications.

Technotomy has since been extending this technology into developing a more generic tool for identifying and measuring web indicators for any type of industry. Their recent merger with another local SME, iVeridis, to create Synoptic Technologies has given access to major clients and enabled the underlying platform to be applied to new areas such as demand and supply management and technology discovery.

