



KB CRAWL 4 AND SPECIALIST MODULES

With KB Crawl Suite, rapidly deploy a range of professional Competitive Intelligence software

KB Crawl 4 is the solution for those who need an **automated Website monitoring service** and intelligence report distribution directly from the intelligence team's workstations. This software suite has been designed principally for information professionals who are looking to **gain time and improve efficiency** compared to manual searches.

KB Crawl Suite is built around the **KB Crawl 4 intelligence engine** which is the heart of this technology. KB Crawl Suite manages the intelligence process from data collection, to processing through to the distribution of information.

Additional modules can be integrated to optimise and customise the monitoring process and connect to complex sites, manage email distribution, filter information on specific pages or export data in XML format to other applications.

KB Crawl Suite is a software package based on the KB Crawl 4 intelligence engine, to which the following specialist modules can be added as desired:

- **KB Scraper:** defines the areas to monitor
- **KB Mail:** distributes and validates emails
- **KB NavRecord:** records browsing operations on complex sites
- **KB Exchange:** imports and exports alerts in XML format
- **KB OCR:** conducts OCR analysis on the textual content of images

ASSISTANCE

Our KB Crawl SAS consultants provide the following services:

- **Installation** and integration of KB Crawl solutions into complex information systems,
- **Technical training** on our solutions and on search and monitoring methods,
- **Maintenance** and **customised technical support**,
- **Advice** and project management for your Competitive and Business Intelligence issues,
- **Follow-up** and leadership of Competitive and Business Intelligence working groups,
- **Sourcing and configuration** for your knowledge bases.



KB CRAWL 4 FUNCTIONS

FUNCTIONS	BENEFITS
Collection	Monitor the visible and invisible Web (HTTP and HTTPS), discussion forums , newsgroups (NNTP), blogs, RSS feeds, newsletters , FTP sites and Intranets. Collect and index documents in the following formats: HTML, RSS, XML, Word, Excel, PPT, Flash, TXT, JavaScripts . Monitor non-Latin languages (Arabic, Mandarin, etc.). Track more than 100,000 pages per day .
Filtering	User-defined alerts according to changes in page content , appearance of key words , using complex queries (KB NavRecord, Boolean operators, truncations, masks) etc. Filter URL addresses or specific areas on each page (KB Scraper).
Processing	Create a filing structure for sources and key words. Schedule the monitoring process on a weekly basis, down to every minute. Index the textual content of images (KB OCR).
Capitalisation	Archive different versions of documents. Benefit from the new search engine incorporated in the application. Option to share a knowledge base.
Distribution	Generate intelligence reports according to user-defined alerts and recipient list. Manage the email distribution process (KB Mail). Create XML format exports for other applications (KB Exchange) or to office documents (KB Scraper).
Interface with other solutions	Option to export KB Crawl Suite data to search engines, mapping, Text Mining or semantic analysis tools, to an intranet or collaborative portal.

TECHNICAL CONFIGURATION

CLIENT	SERVER
Memory: 1 Go RAM Processor: Pentium IV / 2 Ghz or more Operating system: Microsoft Windows XP Pack SP2/Vista Browser: Microsoft Internet Explorer 6.0 min. Hard disk: 5 Go minimum	Memory: 4 Go RAM Operating system: Microsoft Windows Server 2003 or 2008 Other software: SQL Server 2005 and .NET Framework 3.5 Browser: Microsoft Internet Explorer 7 Hard disk: at least 300 Go
Architecture	Single workstation: Simple installation on each workstation. Server: Same knowledge base shared by several intelligence professionals. Crawling, monitoring and email distribution are managed by the server for enhanced performance.
SaaS mode or hosted	SaaS mode frees the organisation from having to install software and enables it to expand and speed up information distribution processes. Hosting on client servers guarantees optimal confidentiality and greater flexibility in terms of configuration.