

## **The Offline Web Archive 200.035.200<sup>[1]</sup>**

The *Offline Web Archive (OWA)* is a system for the selective archiving of web content.

### **Redundancy-free data storage**

With the *Offline Web Archive* internet and intranet content can be crawled and archived in user-defined intervals. Due to the redundancy-free storage of data, the memory requirements are kept to a minimum.

### **Long-term storage**

The *Offline Web Archive* supports the functions for long-term storage as required by ISO 14721. In addition, data can be stored in system-independent "containers" with complete XML-based metadata description (Archival Information Packages – AIPs). AIPs enable lossless data and metadata interchange between the *OWA* systems.

### **Web-wide availability**

The web client offers access to the entire range of functionalities of the system. All features are available web-wide compliant to the respective security policies due to integrated software distribution.

### **Inventory and collection building**

The establishment and management of collections and inventories is simple and can easily be expanded and adjusted to the archive organization and the inventory structure. The administrator can establish a hierarchical management structure with an unlimited number of levels with various projects or project groups.

### **Project-specific crawling parameters**

The crawling parameters and intervals can be adjusted according to the specific criteria of each project. The system manages and monitors the execution of processes.

### **Selective editing and online analysis**

The "online synchronization" feature allows a detailed selection of the pages and files to be archived. The integrated site analyzer enables a file- and protocol-specific selection of the resources of a page.

### **Archiving dynamic multimedia content**

The *OWA*-specific feature for the analysis and storage of dynamic multimedia content ("flashes") allows the archiving of pages with fully functional flashes. Video streams and other dynamic content remain fully functional in the archived version.

### **Easy navigation in the archive**

Regardless of the quantity of the archived web content and documents users can easily research the archived data, read or print archived documents and pages and store archived content on a local storage medium. Navigation in the archive is easy; the system supports various uses such as the reconstruction of complete web sites at different times or the tracking of changes over a freely definable period of time.

### **Full-text data search**

All metadata and primary data can be accessed via a full-text database. The number of hits can be limited by criteria such as inventory categories, date or period of time. Hits will be

displayed either as link lists or flow statistics. The drill-down features enables user-friendly research in the flow statistics.

### **Data and maintenance analysis**

The integrated analysis tools enable detailed analysis of the web content to be archived, from the amount of data to file types and versions. The integrated transformation tool provides options to convert (transform) media and file types or file versions without any loss of the original files.

### **User management**

The integrated user management ensures that only authorized users are allowed to create and edit the projects to be archived.

## System requirements

Last update: January 2012

All functions can be installed and used on one computer system. For optimal performance when processing large amounts of data, especially while running crawlings, at least three server systems should be employed: A user client for all administration and research tasks, a "DataAnalyzer" for processing the crawlings, caching the crawled data and preparing the crawls for being archived and a database server.

### User client

Hardware: Intel Pentium 4 processor or equivalent; 2 GB RAM main memory recommended.  
Software: Windows XP (SP3 or later) or higher; .NET 3.5 SP1, Windows 7.

### DataAnalyzer

Hardware: 1 dual-core processor, 2 dual-core processors or equivalent recommended; at least 4 GB RAM main memory, 8 GB RAM recommended; at least 250 GB disk memory.  
Software: Microsoft Server 2003 or higher; .NET 3.5 SP1.

### Database server and ServiceBroker

Hardware: 1 quad-core processor, 2 quad-core processors or equivalent recommended; at least 8 GB RAM main memory, 16 GB RAM recommended; at least 1000 GB disk memory,<sup>[2]</sup> 2000 GB recommended. Software: Microsoft Server 2003 or higher; NET 3.5 SP1; Microsoft SQL Server 2005, Microsoft SQL Server 2008 recommended.<sup>[3]</sup>

---

<sup>[1]</sup> Last update: March 2012

<sup>[2]</sup> The system parameters for your *Offline Web Archive* vary depending on the quantity and frequency of URLs to be archived. We will assist you in planning your *Offline Web Archive*.

<sup>[3]</sup> Other databases on demand.