



Organize-n-Search

State-of-the-art Low-budget Document Management Solutions

"We are living in the information age... The information explosion..." We have heard it so many times that have stopped paying any attention to it. However, information penetrates into every aspect of our lives. We are constantly trying to acquire new knowledge and looking for opportunities to benefit from it.

Users who actively work with documents and information, frequently face the problems related to search, organization and efficient use of documents. Copyeditors, writers, journalists, researchers, analysts, consultants, lawyers, medical workers, students, all run into the same challenges at home and at work.

This paper is intended for a wide range of people, who, for personal or business need, work with a large number of documents and other information. We take a close look at the problems of information management, benefits of using advanced technologies in the low-budget personal information management system, as well as system selection criteria to meet personal and professional needs of information workers.

Challenges of Document Management

Recent developments in the fields of computer systems and data storage have led to a new question: how can we effectively manage digital information?

Nowadays big part of information is stored in a form of text: books, articles, reports, memo, notes, specifications, descriptions, whitepapers, and manuals, not to mention a huge amount of time sensitive information, such as invoices, bank statements, schedules, contracts, and tax returns.

Yesterday, papers, photo albums, music disks, and video tapes were kept in drawers, boxes, and cabinets. But the development of personal computers and Internet has started the era of digital information.

Development of electronic formats has significantly increased system storage capacity and allowed accumulation of large information volumes. However, recent developments in the fields of computer systems and data storage have led to a new question: how can we effectively manage digital information?



Recent studies by IDC¹ revealed that on average 13 hours of every 40-hour work week are spent on creating documents. 9.5 hours per week are spent on searching for information, while almost 9.6 hours on analyzing the information. 6.5 hours are wasted on searching for information that is never found leading to the need to recreate the content. Formatting of information between different applications takes about 3.8 hours per week, whereas version control related issues take 2.2 hours.

Issues, effects and implications of information management are summarized in the following Figure.

Issues	Effects	Implications
Slow search Search without desired results Redundant search Recreation of documents Difficulty of use of the found information	<i>Employer</i>	
	Unplanned for wasted time	Missed deadlines Project failure Lost revenue Loss of employee
	Work slowdown	
	Decrease in productivity	
Decline in quality		
	<i>Employee</i>	
	Increased workload	
	Negative attitude towards work	
	Decline in the level of satisfaction from the job	

Figure 1: Issues, effects and implications of information management

- What is the best way to organize the information to find it faster in the future?
- How to easily find information inside of large volume of materials?
- How to find documents that are related?
- How to save the search results and view them in the future?
- How to share found information with colleagues and friends?
- How to effectively use found information?

Importance and significance of those problems are major factors that stimulate the development of new solutions and information management systems. Information Retrieval, Data and Knowledge Bases, Document & Content Management, to name a few, are the branches of information technologies that deal with the problems of information management.

¹ Susan Feldman, Joshua Duhl, Julie Rahal Marobella, Alison Crawford. The Hidden Costs of Information Work. March 2005



Solutions to problems related to document management are tightly linked to the following challenges: improving the efficiency of access to information, improving quality and speed of the search, improving the efficiency of information processing, and improving reliability and safety of storage

Best solutions for effective search of information are based on search engines and information retrieval technologies

Solutions to Document Management Problems

Solutions to document management problems are tightly linked to the following challenges: improving the efficiency of information access, improving quality and speed of search, improving the efficiency of information processing, improving reliability and safety of storage.

Efficient Access to Information

It is necessary to quickly and easily extract the text documents which meet certain criteria from an array of available information. These requirements are diverse and constantly changing. For example, original sources for articles, data for reports, textbooks to prepare for the exam, patient's medical records, or precedents for court case – all have high, but temporary value to resolve the pressing challenges.

After finding the required documents, working through them, and creating a number of versions, the user will need to consolidate and store the results. For example, one may need to save a set of documents, or add comments to a set of documents for future use. One possible solution to meet the changing needs is to place a document in several groups. A group could consist of documents on certain topic, papers of the same author, articles of the same journal issue, previous versions of the article, or materials used to write an article.

Searching and organizing information in a meaningful way takes up a lot of time. To shorten the cycle and make a process more enjoyable, a number of solutions have been proposed.

Quality and Speed of Search

In some cases users can find the documents they need by using a query – a word or combination of words that might be in those documents.

In the past, search required scanning of all files on the computer drives and going through their content comparing the key words with words in the document. This called for the sequential scanning of all files for each request. But increased size and number of files have dramatically slowed down the search process. In addition, morphology was neglected and multiple queries were needed to find the document.



Best solutions for effective search of information are based on search engines and information retrieval technologies. The entire collection of files is pre-processed and the information about the documents and key words is stored in the index files. Indexing works for various file formats and takes into account all possible forms of the same word. This “smart” pre-processing mechanism significantly accelerates the search and improves its quality.

Organization

Top notch tools to organize the information use multiple hierarchical categorizations which came from the domain of knowledge bases and ontologies

In many cases the user is unaware of the words contained in the document of interest. It’s also possible that the user is not able to generate a query that returns desired outcomes, or the number of documents is too large, or some documents may not contain the right words. In these scenarios the user has no choice but manually look for a desired document. To save the results of manual search, many use the systems designed specifically for organizing the information.

Simplified versions of organization systems use fields and registration cards to link the documents and accompanying information (date, author, title, a brief description, etc.) However, field sets are fixed and limited, and often do not allow grouping of the documents to accommodate changing needs of the users.

Enhanced systems use a hierarchy of folders (catalogs, or directories). However, in most cases, when a document belongs to multiple topics, the user may end up facing several problems. For example, in the hierarchy of file system folders, a document can not be assigned to several folders without duplication. In this case, duplication may result in an unnecessary increase of information volume as well as inconsistencies in content after one of the documents has been modified.

Top notch tools to organize the information use multiple hierarchical categorizations which came from the domain of knowledge bases and ontologies.

Modern document management systems include version control features that allow access to the previous versions of the document and history of changes

Version Control

Authoring of a complex document is a long process and requires many edits, corrections and rewritings. To avoid confusion, it is necessary to maintain a history of changes in the document. The old-fashion solution was to save the changes in the separate file with a unique name, which



often resulted in lost files, more storage space as well as difficulties in finding the right version of the document. These and other problems related to tracking the history of the content, storing different versions of the document, and returning to its previous versions have been addressed by the invention of the versioning systems. These systems are designed to provide access to the previous versions and history of changes.

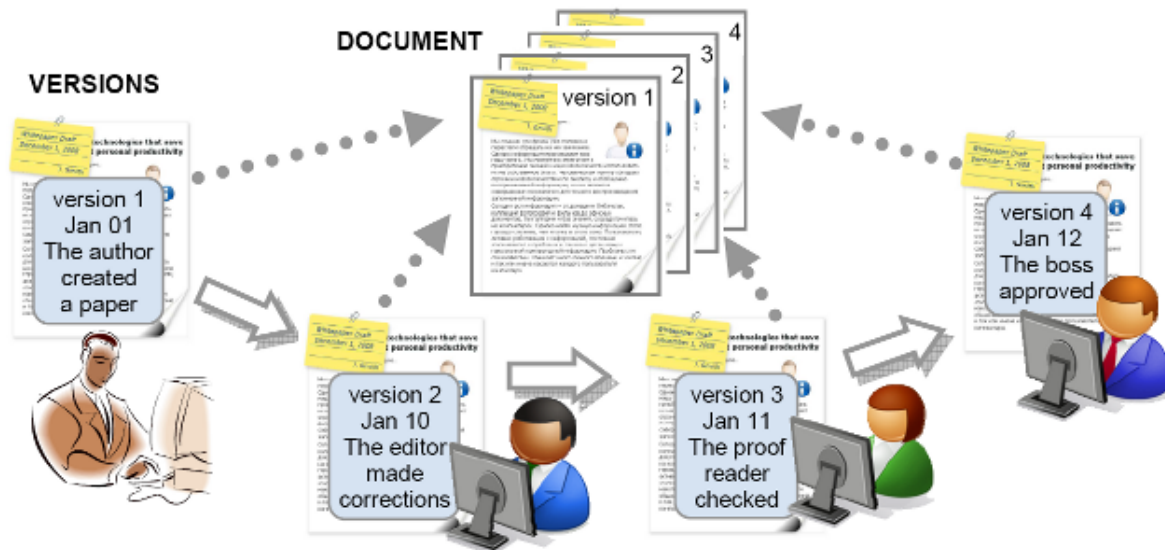


Figure 2: Authoring a document

Effective Work with Information

To eliminate unnecessary labor and reduce the amount of wasted time, one needs an integrated solution that combines search, editing and version control functionality

Search, organization, and version control, by themselves, significantly simplify the process. But till now, most of these functions were only provided by separate software tools. The first program implements search. The second program organizes information. The third program edits it. The fourth program keeps version history. And so on.

A user has to run multiple applications, toggle between them, import and export documents, and move and copy the files. This process dramatically slows down the work, decreases productivity, increases pressure, and therefore leads to mistakes and reduces work satisfaction.

To eliminate unnecessary labor and reduce the amount of wasted time, one needs an integrated solution that combines search, editing and version control functionality.



Privacy, Security and Reliability of Storage

It goes without saying that information is a valuable resource that is expensive to produce. It is necessary to not only provide a safe storage for the entire set of documents, but also protect valuable information from computer hardware and software failures, as well as human errors. In addition, the confidentiality of information should be preserved – unauthorized users should not have access to the information without the permissions from the owner. However, if necessary, the results of the work have to be publishable to third parties.

Earlier applications stored files on the secure computers in a folder structure. Individual users had access to specific folders, which required a complex access rights management policy. Thus the information was often duplicated on the users' computers, causing many problems related to information relevance.

To address the above mentioned problems, modern document management systems store information in centralized repositories, which make it easy to store, retrieve, manipulate and modify documents. Advanced repositories support storage and processing of multiple documents and file formats including, but not limited to: text (Word, Acrobat, Open Office, etc.), spreadsheet, fax, e-mail, audio, and images.

Documents, images and other information stored in the electronic repository are easily accessible and retrievable. The losses associated with errors in streamlining, organizing, and placing of the documents are drastically reduced and possibly even eliminated. In addition, each document keeps not only a history of who viewed it, made changes and what changes were made, but also other information about the document, such as title, contents, themes, etc.

Valuable Benefits of Document Management Systems

Thus, state-of-the-art information and document management systems

- reduce information processing time (multi-category systems allow for fast categorization of the incoming information and re-organization of existing information)
- reduce the time required to access the information (full-text search tools and category system, history and version control provide an easy and quick way to find information)

Modern document management systems store information in centralized repositories, which make it easy to store, retrieve, manipulate, and modify documents



Document management systems reduce the time spent on creating, accessing, and processing a document

- reduce the time required to create a document (integration of search, organization, modification and version control features in a single platform allow the user to work on new and existing documents in a more effective manner)
- eliminate the cases of lost data (electronic repositories automatically capture all document changes and allow the user to restore the history of changes)

By leveraging a wide range of features provided by information management tools, one may free up the time normally spent on unnecessary tasks and focus on more important activities. As a result, the use of information management systems increases the quality of work.

Criteria for Selecting the Right Document and Information Management System

Criteria for selecting the right document management system:

- *Flexible categorization*
- *Powerful search tools*
- *Central repository*
- *Composite documents*
- *Document registration cards*
- *Supported file types*
- *Versioning system*
- *Navigation history*
- *Easy-to-use interface*
- *Modern technology and architecture*
- *Integrated solution*
- *Low cost of ownership*

Flexible categorization: The system must support the categorization of documents to meet specific requirements of the user. To do that, the system should include the following features:

- Flexible categorization (user should be able to create any categories or topics and place the documents there)
- Hierarchical categorization (high level topics that consist of more specific topics)
- Multiple categorization (the same document might be included in several topics, categories or groups of documents)
- Ability to merge related files in a package

Flexible grouping that keeps the history of the results simplifies future access to documents inside of assigned topics, and allows one to see the relationships between documents found in one category.

Powerful search tools: The system should be able to perform a full-text search of information by query which contains individual terms or phrases. The search feature should

- be fast, which implies indexing
- support full-text search for all common formats – pdf, doc, odt, etc.
- take into an account the differences in spelling of various grammatical forms of the words
- work with individual repositories, categories and themes (topics)

The above mentioned features allow the user to effectively query the documents, provide a fast access to desirable



documents, and make it possible to work on documents that have not yet been classified.

Central repository: The system should be able to store information in a centralized repository that allows:

- storing high volumes of documents
- creation of multiple personal repositories
- protection of confidential information

Documents in the system should not be viewable by other applications. Only the owner of the information should be able to grant the access to the repository.

Repositories not only eliminate the need to manually create the files and directories, but they also restrict access to information, tighten security and improve reliability by providing backup, recovery and data protection tools.

Composite documents: The system should be able to work with the collection of files as a single unit, allowing the user to make changes to the set of documents. This functionality helps to improve usability and makes it easier to work with documents that consist of multiple files – for example, html documents with pictures.

Document registration cards: The system should support the functionality of attaching useful information, such as name, purpose, abstract, comments, author, date of creation and modification, etc. to the document or file. This type of information helps to increase the accessibility of the documents. The information about the document should be flexible enough to adapt to the needs of the user and the information unit type.

Supported file types: The system should be able to support a wide range of common document types and formats, including Microsoft Office (Microsoft Word, Microsoft Excel, etc.), Open Office, as well as the formats of scanned documents and images.

Versioning system: The system should be able to support multiple versions of the document, track history and changes in chronological order – who, when, why modified the document and which changes were made. If needed, this functionality enables the user to work on one of the previous versions of the document.

Navigation history: The system should record the sequence of events describing the steps the user took while working on the documents and have that information available to the user at any given time.

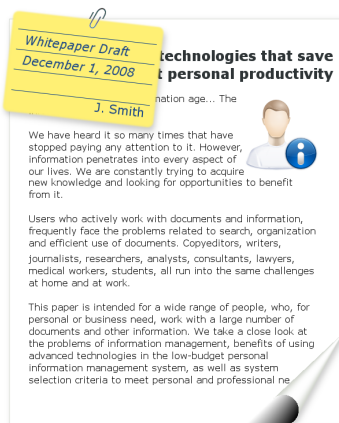


Figure 3: Composite document



The system should have easy to navigate interfaces with panels displaying categories, history, versions, and search results

Easy-to-use interface: The system should provide a user-friendly interface that includes intuitive navigation as well as the panels displaying categories, history, versions, and search results. All of these will dramatically enhance user experience and therefore increase user satisfaction.

Modern technology and open architecture: The system should be built using the latest technologies. The architecture should be

- scalable – support an unlimited number of repositories, documents stored in a repository, categories and their levels, as well as a fast search through unlimited amount of information
- modular and expandable – provide a foundation for rapid development and fast delivery of new features requested by the users
- cross-platform – compatible with Windows, Linux, and MacOS operating systems

This allows the system to grow organically and reduce the time to deliver the new features to meet growing user needs.

Integrated solution: The user's objective is an effective execution of her or his work. To accomplish this goal the user has to go through repetitive cycles of work with information and documents. These cycles may include:

- Gathering of the information for a document
- Analyzing information
- Creating the outline and the first draft of the document
- Placing the document to the repository
- Making changes to the document
- Preparing the document for future use
- Searching for other materials that will be used in a new version of the document

These phases are executed repeatedly to improve the quality of the document, bringing it to the desired results. A good system should be able to integrate the above mentioned features so that the user can complete the sequence of document development tasks in a single system. This implements agile document management.

Low cost of the ownership: Adoption of a document management system can save any organization millions of dollars. At the same time, the scale and broad functionality of corporate systems leads to the high cost of ownership unaffordable for personal users. It's also important to note that a user might not need all the features available in a corporate system and therefore will only get overwhelmed

The cost of a personal information management system should be low. At the same time, it has to provide the right set of features to match the needs of individual user

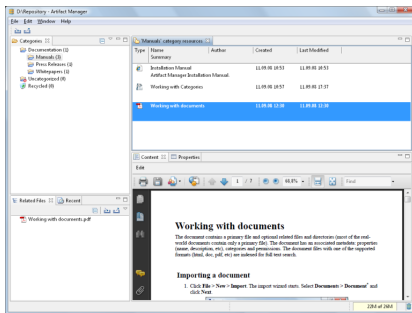


by its complexity. The cost of a personal information management system should be low, but at the same time it has to provide the right set of features to match the needs of individual user. The system should be easy to install and run on any personal computer.

Artifact Manager

Artifact Manager is the first enterprise-class personal platform for document and information management

Artifact Manager is an advanced document and information management system. This simple, convenient, low-budget solution has all of the features of the enterprise information management system that helps to achieve higher productivity levels through a better management of personal documents and information.



Required features	Artifact Manager
Flexible categorization	✓
Powerful search tools	✓
Centralized repository	✓
Composite documents	✓
Document metadata	✓
Wide range of file types	✓
Version control	✓
History	✓
User-friendly interface	✓
Modern technology & architecture	✓
Integrated solution	✓
Low-cost ownership	✓

Figure 4: Features of Artifact Manager

Artifact Manager is the first enterprise-class personal platform for document and information management. It combines a powerful search, flexible organization, reliable storage, and convenient interface in a single easy-to-use environment.

Download Artifact Manager *now* at <http://www.ArtifactManager.com/downloads.html>
No obligation of buying, no cumbersome registration, no spam