

White paper

Documill Visual Search – Fast discovery of relevant information

The challenge

An average corporate citizen spends 12 hours per week looking for information – that is 30% of effective work time.

60% of the employees say that search into corporate content repositories needs improvement.

People waste 3.5 hours each week on searches which don't turn up the right information.

Employees, managers and business owners need to have all the relevant information available when making business critical decisions. This information is often scattered across shared file directories and content repositories. The performance of information dependent worker is far from optimal when most of the time is spend on finding the information, not using it.

Unfortunately, quite typically there are no simple methods to access all this multi-page formatted, visual information. As a result, employees spend an average of 12 hours per week looking for information, and even more time recreating information content that already exists. This creates a sense of frustration on the part of employees, leads to severe losses of productivity and uninformed decision making. Not surprisingly, 60% of employees say that search on their company's content repositories needs improvement and half of the information they find during searches is useless.

Research by IDC suggests that Information workers spend 9 – 10 hours a week looking for information. They waste about 3.5 hours each week on searches which don't turn up the right information. So an average company that employs 1000 information workers can expect to waste at least 3,5 M€ in annual salary costs because of time wasted looking for information. Similar type of evidence is also provided by Butler Group who claims that as much as 10% of a company's salary costs are wasted on ineffective searches.

Next to employees, today information is organization's most valuable asset. Getting the right information to the right people at the right time plays an important role in how quickly a company can

respond to emerging opportunities and competitive threats. A single, simple, and relevant search interface to enterprise information is a key component of an information access infrastructure— it helps employees make the most out of company's information by making distributed content repositories easily searchable.

The Visual Search solution

The challenges described in previous section require a search solution designed to work across platforms and applications, a solution that easily adapts in response to new user requirements and integrates seamlessly with your existing information infrastructure. Documill Visual Search offers a new approach to tackle cumbersome task of finding relevant Microsoft Office and PDF documents. A wealth of critical business information is stored in those documents like contracts, manuals, meeting notes, sales and marketing materials etc.

Thumbnail preview provides immediate overview of the search results. It shows only those pages/slides which contain match to user's search query.

Ability to simultaneously view multiple documents instead of just one makes result relevancy verification very fast. Highly accurate, instantly zoomable previews ensure premium user experience.



Figure 1. Screenshot of Documill Visual Search

With Documill Visual Search traditional search engine results are enhanced with graphically accurate presentation of the content. Our solution renders and paginates Microsoft Office and PDF documents. It offers highlighted previews of the pages that match to the search query, coupled with the opportunity to instantly view the whole document with a browser. Quite simply, it improves search result accuracy from document-level to page level. Documill Visual Search introduces new user experience paradigm to discover multi-page document content.

Supported File Formats

There are over 400 million documents available in Internet. Over 95% of them are either PDF or Microsoft Office formats.

To serve the majority of key use cases, Documill has opted to support the most popular multi-page file formats with high accuracy visual preview capability. For example over 400 million multi-page documents are available in Internet, and 95% of them are either PDFs or formats produced by Microsoft Office productivity tools. Some of these documents are stored with older legacy formats, and to meet this need Documill Visual Search currently supports Microsoft Word from version 6.0, PowerPoint from version 95, partial Excel features fit for Visual Search use as well as static Adobe PDF features up to version 1.6. Our product architecture is extensible and can easily support additional file formats if the need arises from our partners or customers.

Supported Content Sources

The Documill Visual Search leverages external search engine index. The solution sits between end-user and the search engine, managing and enhancing the user experience for discovering multi-page documents. The solution has been designed from the ground up to take full advantage of your existing search engine infrastructure. It exploits all capabilities your search engine provides, prior retrieving and processing the targeted documents.

All the biggest enterprise search engines offer APIs thru which Documill Visual Search can leverage existing document indexes and search functionalities.

It retrieves documents from any of the content sources crawled by the search engine, i.e. file systems, SharePoint document libraries, Lotus Notes, IBM CommonStore or EMC Documentum. The solution can be integrated to any search engine. Out-of-the-box there are API implementations for leading enterprise search solutions such as

FAST, Microsoft SharePoint, IBM Omnifind, Autonomy and open-source variants based on Lucene.

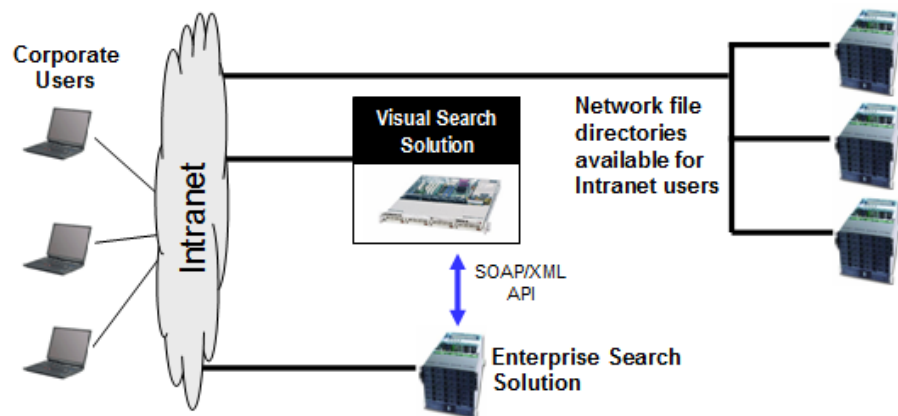


Figure 2. Visual Search as Intranet solution

Besides supporting Intranet needs, Documill Visual Search can be integrated to publicly available Internet indexes provided by leading search providers such as Yahoo and Microsoft Live. Out-of-the-box come our powerful SOAP and XML API support to keep the integration effort minimal into these external search engines.

Premium User Experience

Documill Visual Search user interface has been implemented as a set of customizable web components, complying with the good user interface design practices. Extending and enhancing existing search experience is easy by integrating Documill Visual Search functionality components into your existing search UI. Specifically, you can customize the search result layout using the CSS, XSLT templates or Java scripting. Documill provides you full flexibility in user interface design in order to provide the best possible, consistent user experience for your end-users.

Flexible Security Model

The integrity and protection of the data access is guaranteed through numerous restrictive gatekeepers. Visual Search assumes that the first gatekeeper is always the search engine. The user is first authenticated against the search engine in order to provide only such search results the user is authorized to view. Documill Visual

Search server accesses the content repositories only with valid user credentials. The server supports single sign-on methodologies like use of SSO cookies.

Maximum Scalability

The scalability of Documill Visual Search solution is directly linked to the scalability of Java and available server hardware. In particular, the number of concurrent and total users supported is a function of the web front-ends and available CPU cores Visual Search Preview functionality can use. The document processing is designed to be scaled up just by adding parallel processing power. Thus the system scales almost linearly with more CPU cores. Documill Visual Search fully supports multi-tiered architecture for high-volume scalability thru load balancing and data caching.

Documill Visual Search offers various different deployment models for corporate users. Hosted, SaaS or customer on-site server deployments can be done while enjoying the flexibility for choosing scalable and affordable hardware platform.

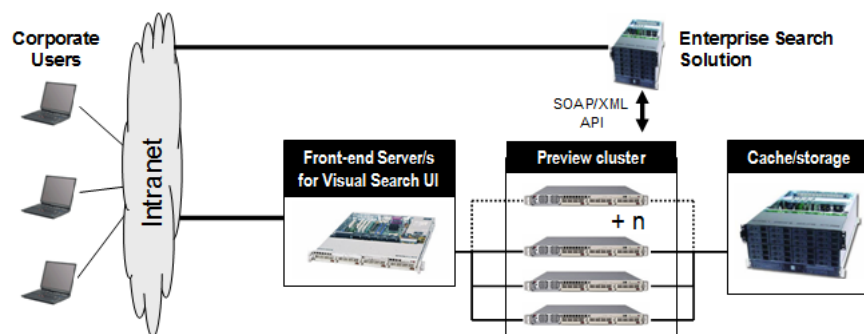


Figure 3. Documill Visual Search with 3-tier architecture

Visual Search Preview Cluster can leverage Smart Caching approach to store results in centralized file storage to minimize response times.

Capacity planning for Visual Search solution is the art of optimizing the hardware (and software licenses) for your particular needs. Finding the right hardware configuration for an expected usage scenario and workload requires understanding usage profiles, user volumes and searchable content characteristics well.

When a product is as flexible and configurable as Documill Visual Search it offers great alternatives for optimization because there are plenty of knobs to turn and variables to configure for those wanting to control every angle of their Visual Search deployment. We do provide reference configurations modeled against typical use cases, but as always with highly configurable and flexible solutions, there is no single “correct” hardware configuration.

About Documill

Documill is an independent software vendor (ISV) offering software solutions for enabling Visual Search, browser-based access to Microsoft Office and PDF documents and server-side content processing for dynamic filtering and ad insertion needs.

With core competencies in server-side enterprise document processing, Documill offers its software solutions to enhance the functionality and user experience offered by existing Enterprise Search (ES) and Enterprise Content Management (ECM) solutions.

Documill's core personnel has worked together in various software ventures since mid nineties, forming one of the most experienced team of digital media innovators in Finland. Coupled with globally unique enterprise document processing capabilities, Documill actively partners with enterprise search players like Microsoft and IBM.

Documill is headquartered in Finland, one of the most technologically advanced countries in the world.