



Empowering
Visual Search

Documill Visual Search Benefits Summary

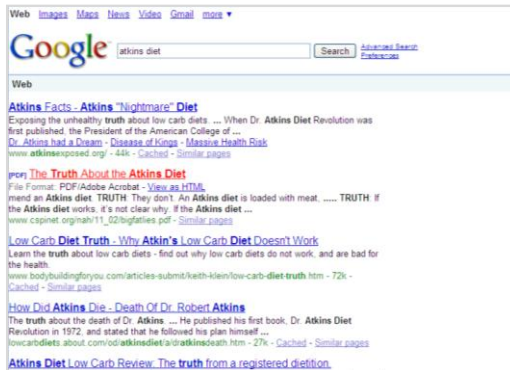
2010-12-07

Disclaimer: All information included in this presentation is company confidential, and should not be redistributed in any media or form without prior written permission from authorized Documill personnel. © 1996–2010 Documill. All rights reserved.

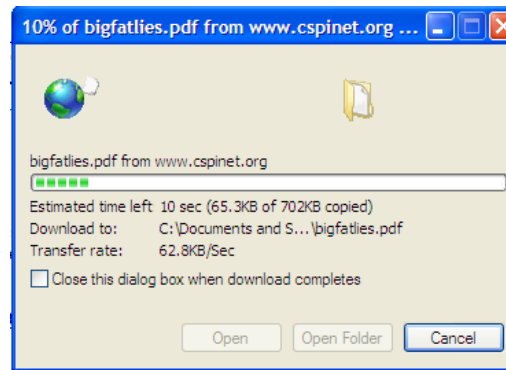
Searching Multi-page Documents is Frustrating

Typically created with MS Office tools,
and stored in editable MSOffice format or in PDF

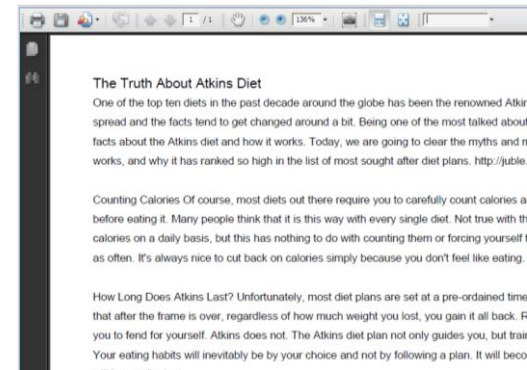
Search



Download



Search again



Ever been frustrated to find the right slides or pages from existing material to a presentation or a document you are putting together?

Documents are everywhere !

Documents are the most valuable asset enterprises create today. Document screening and viewing form the core of any document discovery process.

- Horizontal search
 - Intra/extra/internet
- e-discovery
 - Evidence material for civil or criminal legal cases
- Archive search
 - E-mail attachments
 - PDF-reports from SAP and other ERP solutions
- Forms search & review
 - Public services like IRS
 - e-Office documents
- Cross industry solution
 - Financial services
 - Legal services
 - Professional services
 - Government & Defence
 - Manufacturing & Engineering
 - Medical & Health Care
 - Education & Academic Research
 - Utilities
 - Telecom
 - Retail & Distribution

Enhance Your Enterprise Search Experience!

- Documill Visual Search is a complementary solution running on top of existing content indexing and search capabilities.
 - It serves the audience which main task is to carry out extensive document screening, filtering and search activities.
- Enterprises' most valuable assets are hidden in multi-page documents – and thus documents should be main focus for enterprise search solutions.
 - Focus is on improving discoverability of multi-page documents (PDF, MSOffice).
 - Focus on serving employees tasked to carry out data mining activities.
 - Significantly speed up initial screening efforts where people need to be able to quickly judge the relevancy of a certain document in a context of a given document discovery mission.
- Offers new differentiators and value drivers for enterprise search solutions.
 - Highly accurate page-level document previews.
 - Superior end-user experience with high fidelity, zoomable content previews.
 - Multiple documents previewed and screened at once, without having to download them.

Ensure Critical Documents are Found!

- Documill Visual Search ensures that your preferred search solution has the best tools available for finding and accessing documents.
 - Significantly speed up content discovery experience.
 - Focus on most valuable corporate assets; data hiding in multi-page documents (MSOffice, PDF)
 - Enable high-fidelity content viewing without expensive Microsoft licenses
- Your users may already rely on search functionality, so let's make it easier for them to:
 - Verify search result relevancy
 - Improve insight into documents; page-level, document length
 - Datamine existing documents by browsing recognizable, visual characteristics
 - Offer fastest access into content (shortest time it takes to view first pages)
 - Enable visually enriched, enjoyable user experience (UE)

Visual Search User Experience

No need to have content specific viewers/tools to discover documents.
Selected documents can be consumed without downloading and opening them.

Thumbnail view provides immediate overview of the search results

Filter only pages with matching content visible

Simultaneous view of multiple documents instead of just one – faster relevancy verification

Instantly zoomable page previews from thumbnails

Keywords are highlighted even in the thumbnails

Ribbon view for understanding page context and quick preview of the whole document

Documill Visual Search Demo, using Microsoft Bing API 2.0

dreamliner

.pdf .ppt .doc

Results 1 - 30

1. The Boeing 787 Dreamliner: More Than an Airplane
The 787 is a Complete, Flexible, Efficient Family 787-8 223 passengers (three-class) 8,500 nmi / 15,700 km 787-9 259 passengers (three-class) 8,300 nmi / 15,300 km 787-10 300 passengers (three-class) 8,500 nmi / 15,700 km

2. Dreamliner rolls out
February 2007 BOEING FRONTIERS ** n COMMERCIAL AIRPLANES B y a daM M organ A s airlines await the arrival of the 787 Dreamliner. 787 emp... and media witnessed a first ...

3. Designing an airplane interior is important. The 787 Dreamliner ...
Designing an airplane interior is important. The 787 Dreamliner Catalog "Poss ibilities" presents a sample of the 787 Dreamliner selectable int...

The interface shows a search bar with 'dreamliner', a filter for file types (.pdf, .ppt, .doc), and a grid of document thumbnails. A ribbon view is visible at the bottom of the thumbnails. Blue callout boxes highlight specific features: 'Thumbnail view provides immediate overview of the search results', 'Filter only pages with matching content visible', 'Simultaneous view of multiple documents instead of just one – faster relevancy verification', 'Instantly zoomable page previews from thumbnails', 'Keywords are highlighted even in the thumbnails', and 'Ribbon view for understanding page context and quick preview of the whole document'.

Measurable Benefits

- **Faster access to the content**
 - Makes documents with visual content first class citizens
- **Bandwidth savings**
 - Very relevant for distributed, global organizations (intranet) with varying network topologies
- **Improved productivity**
 - Faster, more accurate discovery
 - Enable re-use of existing document assets
- **Better security**
 - Access documents without downloading them into client side
 - Verify document relevancy prior downloading them – less documents downloaded into browser cache unprotected.
 - Secure way to view documents
 - Without client-side exposure to possibly malicious content (viruses, worms)

Documill Visual Search Benefits

- Speed up the document discovery process → 5-7x gain
 - No need to download documents for screening
 - No need to carry out secondary searches within document viewers
 - Search and visually screen thru larger amount of data
 - Preserve bandwidth while enjoying better and faster user experience
- According to IDC, Information Worker spends 9.5 hours a week for searching information within enterprise
 - Most of the information is in documents spread around different document repositories or file shares
 - Documill estimates that with significant gains in productivity users can save up to 3 hours/week when seeking information hidden in enterprise documents
- The network bandwidth saving can be significant for search related content downloading activities
 - Visual Search provides immediate view of the content relevancy without forcing user to download the document
 - Avoiding unnecessary download can save up to 90% of bandwidth per document

Economic Way to Access Content

- Traditionally, search engines extract text from documents to create a single index entry.
 - This index entry describes the document.
 - Regardless of the size (amount of text, length in pages), each document gets its own index entry.
 - The search accuracy is always "document-level", requiring users to
 - download the document
 - review it using 3rd party viewer
 - perhaps carry out secondary searches within the document to find a relevant piece of content located in some page/s within a document.

...and repeat the above for each document screened.
 - As Documill Visual Search offers page previews of the document, it can quickly pinpoint those pages which contain match to the search query.
 - User can quickly view a certain page, without downloading the whole document.
 - Provides better visibility into the document, promoting the logical (page-level) division of content within a document.
- This results dramatic speed-up of document discovery, enabling users to quickly recognize documents and/or specific pages.

Economic also Bandwidth-wise

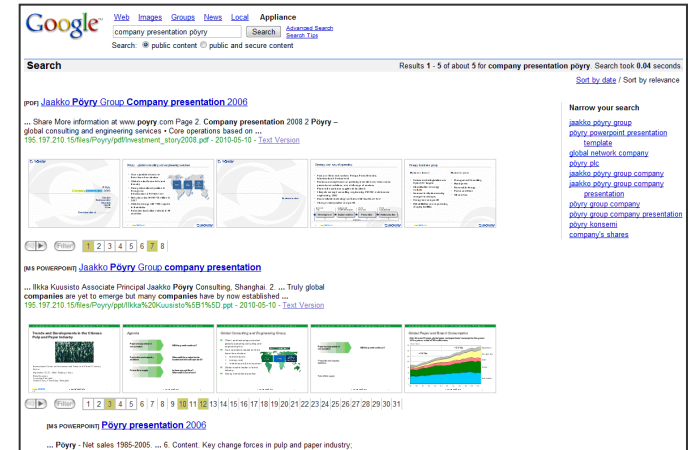
- Visual Search approach allows verifying the relevancy of the document prior download it.
 - Typically, the relevancy check takes about 10 % of the downloading bandwidth.
 - For example - if you have a result set of five documents where only one is relevant for you to download, your bandwidth saving would be 70 %.
- With the above logic applied, your bandwidth savings could be:

Relevancy ratio	Bandwidth saving -%
one-out-of-one	-10
one-out-of-two	40
one-out-of-three	57
one-out-of-four	65
one-out-of-five	70
one-out-of-six	73
one-out-of-seven	76
one-out-of-eight	78
one-out-of-nine	79
one-out-of-ten	80
- Relevancy ratio above means how many documents you need to screen to find what you are looking for.

Documill Core Functionality

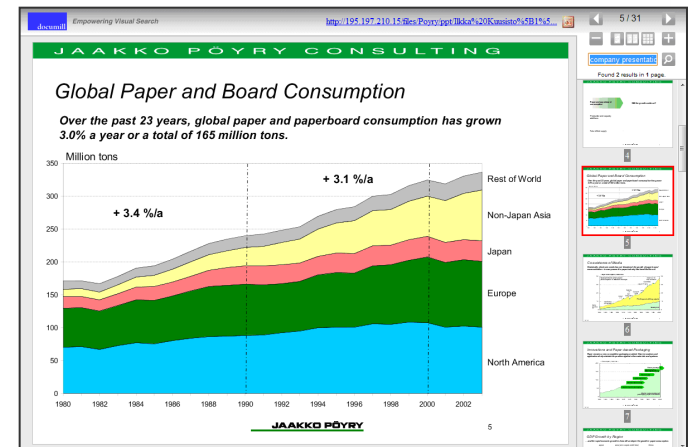
1. Search Resultset Previewer

- See what you have found without having to download those documents
- Get a good overview of your findings prior drilling into those which look most promising
- Fastest way to find those pages which matter most



2. Document Viewer

- No need to download and open documents with some 3rd party viewer; use browser instead
- Most efficient way into the content
- Secure way to view documents which might contain harmful payloads



About Documill

- Documill is an independent software vendor (ISV) enabling browser-based access to MSOffice and PDF documents and server-side content processing solutions.
- Documill's core personnel has worked together in various digital media ventures over ten years.
- Our core competency is server-side enterprise document processing.
- We are located at Innopoli 2:
Tekniikantie 14
FI - 02151 Espoo
FINLAND - EUROPE

