
Compound Documents: Google May Be in Your Future

On a recent flight from Europe to the United States, I had nine hours in which to ponder Google's most recent announcement about Google Docs. I learned that Google Docs supports video playback in a memo or product specification draft.

If you have been a keen student of Google, you may recall that inserting YouTube videos in Google documents has been supported for several years. I recall watching a 2008 "how to" video on YouTube.com The program was created by Labnol.org, and when I checked at the end of January 2011, the program was available from "How to Embed YouTube Videos in Google Docs." <http://www.labnol.org/internet/office/embed-youtube-video-google-docs-presentations/2769/>. Google also provides helpful information in its own Google Docs help. You can access this information at [this link](#) or click <http://docs.google.com/support/bin/answer.py?hl=en&answer=94194>. The Google Groups thread from 2008 contains some conflicting information with one person stating "What is a fact is that you can *not* embed videos in Google Docs documents." http://groups.google.com/group/How-to-Documents/browse_thread/thread/f0d967c6ae68bb96

On January 6, 2011, Google announced in the Docs Blog:

*Happy New Year! We're starting this year by making it easier to view memories from 2010 that you've uploaded to your document list directly in your browser. Just click on a video in a [supported format](#) and press play. Please note that some videos uploaded earlier last year might not have been processed yet and will be available soon. Also, newly uploaded videos may take some time to process before they are available for viewing. So what are you waiting for?
<http://googledocs.blogspot.com/2011/01/this-week-in-docs-video-player-in.html>*

Good question. But Google's question triggered three questions that I want to explore in this month's column.

1. Will the content of the video be searchable?
2. Can organizations with search systems from mainstream vendors of enterprise search index video content?
3. What are the implications of Google's new push for compound documents?

The first question "Will the video content placed in a Google document be searchable?" is, "Probably". Google continues to enhance the search functionality of its video services. In the official Google announcement, the reference to "processed" signaled me that Google is

setting up a pipeline system so that video content in a Google Docs document can be searched.

If you have not explored Google's enhanced YouTube.com search, you can try it. Navigate to www.youtube.com and enter the phrase "Arnold Exalead". You will get two hits to videos in which I appeared as "talent". What is indexed are the metadata attached to the video, not the content of the video itself. Google's interest in video picked up in 2006, a year which marked an uptick in the number of technical papers and patent documents referencing video. After buying YouTube.com in 2005, Google was in rich media mode. Google's effort to index the content of individual videos surfaces in a number of interesting patent documents. I found the patent application US20100008547, "Method and System for Automated Annotation of Person in Video Content" suggestive. Google's engineers were divulging a way to identify people in videos. At the same time, Exalead—a company I think of as the European equivalent to Google in engineering expertise—debuted its Voxalead technology. You can see a demonstration of Exalead's indexing the spoken information in a video, at <http://labs.exalead.com/experiments/voxalead.html>. Exalead's approach remains a step ahead of Google's as I write this. I want to return to the implications of indexing the content of an embedded video in the wrap up.

The second question "Can organizations with search systems from mainstream vendors of enterprise search index video content?" The short answer is, "Maybe". Autonomy, for example, offers its video processing component as part of the basic system. The answer is that most of the mainstream vendors support video indexing and search; however, a licensee may need to make provisions to acquire the components required to handle video. Autonomy makes this statement on its Web site

<http://promote.autonomy.com/components/pagenext.jsp?topic=PRODUCT::VIRAGE>

Powered by [IDOL](#), Autonomy Virage Video Search delivers powerful next-generation search capabilities. Along with traditional ways of searching video content (e.g. by channel, program or time), users can reach right inside video streams, navigate vast quantities of rich media content and search by a range of parameters including audio, scene, speaker, location, key frame, on-screen text, face, token and concept.

Exalead and Autonomy are roughly equivalent. To make other mainstream enterprise search systems process and make audio and video content searchable, a mix of third-party products and services may be required. For example, one can use a Microsoft SharePoint and Fast Search environment and integrated technology from Raytheon BBN. Information about the Raytheon BBN rich media technologies is available at this link

http://www.bbn.com/technology/speech/information_extraction_from_text.

The third question, "What are the implications of Google's new push for compound documents?" is the most difficult to answer. Google has been making its consumer-facing products and services more and more robust. Without much pre-marketing, Google then moves an advertiser-subsidized or low-cost service into the Google Enterprise arena. Examples range from Gmail which has been enhanced with close integration to Google's

voice and chat services. In addition, Google makes the Postini security and spoliation prevention services available to enterprise customers. My hunch is that the improved support for video in Google Docs is a precursor to increased functionality. Google's history in the enterprise is to put significant pricing pressure on incumbents. As a result, I think Google will incrementally improve its video search within its enterprise offerings. Once a customer base is in place, Google will build out its offerings on what will be a *de facto* compound document platform.

The interest in rich media has surprised me. I find that Google makes much of its technical information available via YouTube.com videos. When I meet with a start up with a strong contingent of 20-somethings, informative videos take precedence over more traditional modes of communication. If the start up must decide between a 10 page white paper and a video, in the last 18 months, videos are the hands down winner. As more rich media oriented people enter the workforce, the notion of a document will shift from text to a compound document. The "compound document" will contain text, possibly audio, a collaboration function and a record of the comments made in the collaborative space, and other information objects. In short, what used to be a 20 page business plan may become a dataspace stuffed with different types of content and integrated functions that allow a reviewer of the "compound document" to access a history of the changes to the content in that particular dataspace.

Vendors of search and content processing systems cannot rest on their considerable technological achievements. In my experience, making a search system designed to index text and permit key word queries is going to be a difficult—perhaps *impossible* is a better word—job. The companies mentioned as leaders in search, therefore, will have to invest either in research and development or acquisitions to be able to meet customer demands.

Cloud-centric solutions like those from Amazon, Google, and Rackspace may have an advantage. Many organizations have traditionally resisted hosted search solutions. The file size of video content and the required computer processing may be too much for many organizations. As a result, the cloud-based services may have an advantage over on-premises enterprise systems able to make "compound documents" findable. Browsing a number of videos is not feasible. Search systems will have to implement Autonomy- and Exalead-type functions that "jump" to the point in the video where the needed content is located. Most organizations' information technology departments are stretched thin and may lack the specialized expertise required to deal with the challenges of "compound document" processing.

Users—particularly recent college graduates—arrive with a method of working that blends collaboration, real-time information exchange, rich media, and traditional content. The organization hiring these individuals, therefore, faces a number of management challenges. These include implementing information policies that do not drive away the new hires and providing "finding" tools that allow employees to get on with their jobs, not relearn how to perform certain basic tasks such as locating a needed fact.

So search vendors, customers, and users are going to find themselves entering a world in which the traditional notion of a document is changing and quickly. In this new environment, what will Google offer. ? As I write this, Google has changed its top management set up and

it continues to diversify into new lines of business. Based on the information crossing my desk, Google's mobile activities attract considerable attention but have not yet turned into the financial bonanza some stakeholders expect. I think it would be a mistake to see Google's mobile ambitions as distracting the company from its goal of expanding its enterprise business. But Apple is thriving in mobile and its iPad is nosing into many organizations.

Google's mobile devices will perform searches. When a large organization "goes Google", the Google ecosystem may be well positioned to put considerable competitive pressure on today's leaders in enterprise search. Will Google become the hot enterprise vendor in 2012? Microsoft and Oracle will certainly want to keep Google outside of a very cozy tent. But Google's nose under the tent may bump into the snouts of other companies pushing into the enterprise. Some jostling and shoving are inevitable. Excitement is a 100 percent certainty for 2011.

Stephen E Arnold, January 28, 2011

Mr. Arnold is a consultant. More information about his practice is available at www.arnoldit.com and in his Web log at www.arnoldit.com/wordpress.