



JPL Media Search® Project

Multimedia Search Tool for Accessing Engineering 'Best Practice' Information

David Oberhettinger
Office of the Chief Engineer
NASA/Caltech Jet Propulsion
Laboratory

David Tate
Managing Director
Owl Insight LLC

February 26-27, 2008



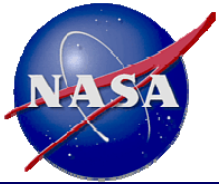
Background



Office of the Chief Engineer

Have you considered videotaping critical engineering meetings and important presentations, but knew you would never sit through 100 hours of videos to retrieve a key fact?

- Multimedia can help NASA effectively capture “best practice” information and forestall the loss of engineering know-how
- But the inability to index and search audio/video files, unlike text files, has limited their usefulness as technical references (or as engineering decision documentation)
- JPL and Owl Insight LLC have developed a system that can index/search collections and find/play the specific video snippet with the search term
- Our pilot-scale, browser-based, search engine can perform context-sensitive searches of audio-video files and graphics metadata, as well as all the text-based file types commonly used in an office environment
- Media Search[®] works by transcribing the audio track to text, performing semantic processing by applying a natural language model to determine the context in which engineering terms are being used, and indexing the information for access via a dedicated browser search page. *Semantic processing* permits a search on the word “gauss” to produce results containing terms like “magnetic contamination” that are not closely related



Semantic Processing



Office of the Chief Engineer

- **Boolean search**: conventional search method that requires you to know exactly the right search terms in order to get successful search results.
- **Mathematical approaches to search**: represent documents as arrays (vectors) of numbers and use neural networks or statistical analyses to match documents with queries. These systems can recognize documents that more or less match the query based on the statistical regularities common to natural language. (That is, based on an assumption that words in a paragraph are not wholly independent of one another.)
- **OrcaTec[®] language modeling approach**: The OrcaTec[®] Information Retrieval Toolkit uses a Bayesian probabilistic model to estimate the likelihood that a document is relevant to a specific query. The software calculates the probability of one word being in a paragraph given the other words in a paragraph. For example, if the word "accelerate" is in a document, it is likely that such words as "force," "engine," or "velocity" will be in the document. Conversely, if the words "force," "engine," and "velocity" are in a document, then it is likely that this document is about "accelerate" even if that word does not happen to be in it. This enables the system to accurately predict which documents that lack the specific search terms are likely to be relevant to the search. The system progressively becomes more knowledgeable— or “learns”— as it is trained on additional documents.



Media Search[®] Demonstration (#1)



Office of the Chief Engineer

Step 1. Online user enters search terms, and selects the file type “Video Files”

JPL Best Practices Media Search Pilot Project - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://jpl.owlinstight.com/jpl/

Frequently Used Links JPL NASA Personal Reference Standards & LL

NASA Jet Propulsion Laboratory California Institute of Technology

BEST PRACTICES MEDIA SEARCH PILOT PROJECT

FIND SEARCH RESULTS THAT CONTAIN:

THIS CONCEPT > cannon launches

THIS EXACT PHRASE >

ALL OF THESE WORDS >

AT LEAST ONE OF THESE WORDS >

NONE OF THESE WORDS >

THIS FILE TYPE >

ALL FILE TYPES
Video files
Audio files
Text

THIS METADATA > Author contains

DISPLAY > 10 RESULTS PER PAGE

CANCEL BEGIN SEARCH

OWL INSIGHT COMPANY CONFIDENTIAL / PROPRIETARY
PILOT-PROJECT DEVELOPED BY OWL INSIGHT LLC UNDER JPL SUBCONTRACT



Media Search[®] Demonstration (#1)

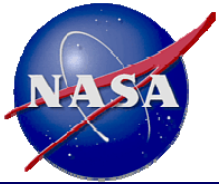


Office of the Chief Engineer

Step 2. User selects a search result [selects “*MER Project: Stealing Success...*” video]

The screenshot shows a Mozilla Firefox browser window with the following elements:

- Browser Title Bar:** JPL Best Practices Media Search Pilot Project - Mozilla Firefox
- Address Bar:** http://jpl.owlinsight.com/search?noField-conceptModifier=%7E%7C0&noField-andModifier=%2B%7C0&noField-phraseModifier=%2B%7C%2B&noField-phrase=&
- Navigation Bar:** File Edit View History Bookmarks Tools Help
- Search Bar:** NEW SEARCH [input field] Start Over
- Search Results:**
 - Found: 1 - 10 of 20 for (+docBody:cannon*5.0 docBody:launches*5.0 docBody:launch*0.45 docBody:data*0.4 docBo
 - Result 1: [A Useful Guide for Chief \(and Other\) Mechanical Engineers - Volume 1 and 2: Part before launches say , aren't you guys were... to launch at or way too much . But t step... as well as going to just collect every piece of data... get on this just part of](#)
01:17:12
 - Result 2: [MER Project: Stealing Success from the Jaws of Failure](#)
there's the more launched Canon launches the parachute is up for shreds is the ye would... and rocket tests and care she says after launch... it to quickly we didn't of
 - Result 3: [A Useful Guide for Chief \(and Other\) Mechanical Engineers - Volume 1 and 2: Part to build the designer launches , What I do well... that on expendable launch miracl would agree which and if you can't find the data and you need some incredible stal emotion amongst the elements](#)
01:27:36
 - Result 4: [A Useful Guide for Chief \(and Other\) Mechanical Engineers - Volume 1 and 2: Part ready to launch but if it doesn't. You'll still have a factor for one euro launch" you l and does... the ahead of us are in clamping down and does... . This is an impact pl](#)
01:27:23
- Image:** A large image of the Eagle Nebula with the JPL logo and the text "SELECT A SEARCH RESULT TO VIEW IT HERE." overlaid.
- Footer:** COMPANY CONFIDENTIAL / PROPRIETARY PILOT PROJECT DEVELOPED BY OWL INSIGHT LLC UNDER JPL SUBCONTRACT




Media Search[®] Demonstration (#1)



Office of the Chief Engineer

Step 3. User plays short snippet from 43 min. video or reads transcript (note misspelling)




Jet Propulsion Laboratory
California Institute of Technology

BEST PRACTICES

MEDIA SEARCH PILOT PROJECT


SEARCH RESULTS:

Found: 1 - 10 of 20 for (+(docBody:cannon*5.0 docBody:launches*5.0 docBody:launch*0.45 docBody:data*0.4 docBo



[A Useful Guide for Chief \(and Other\) Mechanical Engineers - Volume 1 and 2: Part before launches say , aren't you guys were... to launch at or way too much . But th step... as well as going to just collect every piece of data... get on this just part of](#)


01:17:12



[MER Project: Stealing Success from the Jaws of Failure](#)


there's the more launched Canon launches the parachute is up for shreds is the ye would... and rocket tests and care she says after launch... it to quickly we didn't ch

00:43:51



[A Useful Guide for Chief \(and Other\) Mechanical Engineers - Volume 1 and 2: Part to build the designer launches . What I do well... that on expendable launch miracl would agree which and if you can't find the data and you need some incredible stal emotion amongst the elements](#)

01:27:36




[A Useful Guide for Chief \(and Other\) Mechanical Engineers - Volume 1 and 2: Part ready to launch but if it doesn't. You'll still have a factor for one euro launch" you l and does... the ahead of us are in clamping down and does... . This is an impact pl](#)

01:27:23

NEW SEARCH

Start Over

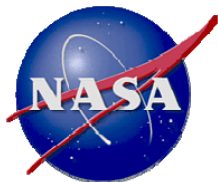
Fun with $\frac{1}{2} \rho v^2 C_d A$



, the suit address an are are empirically direct drive performance and not in an elevator ride so we have to learn so first I was about our in a culture shoe ski shoots his quitters . you see this on TV am sure that that they need a short run this

. in , this is this the this is the drag appear and one half row be squared to be a

bit where first attempt to try to get that drag right here we are were dropping a action tank pressure right there for nothing from helicopters as though there's the more launched [Canon launches](#) the parachute is up for shreds is the year

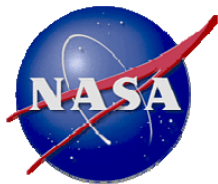


Media Search[®] Demonstration (#2)



Office of the Chief Engineer

Step 1. *All Files* concept search for “gauss” on files that lack the word “gauss”



Media Search[®] Demonstration (#2)



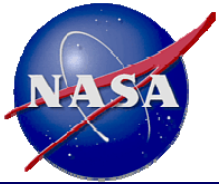
Office of the Chief Engineer

Step 2. These semantic-processed text files contain “magnetic field,” “magnetic susceptibility,” but not “gauss”

The screenshot shows a Mozilla Firefox browser window displaying search results from the JPL Best Practices Media Search Pilot Project. The search query is: `http://jpl.owlinsight.com/search?noField-conceptModifier=%7E%7C&noField-andModifier=%2B%7C&noField-phraseModifier=%2B%7C%2B&noField-phrase=8nc`. The search results are displayed in a list format, showing the following items:

- HTML** (12.10 KB, BEST PRACTICE): Aircraft, Cryogenic Systems, Facilities, Flight Equipment, Ground Operations, Group Materials Processes, Payloads, Pressure Vessels, Range Operations, Risk Management... creating a circular magnetic field around the conductor... of one or more turns a magnetic field lines. [D] Figure 2. Magnetized bars showing directions of magnetic magnetic field and the current increased while adding... the desired magnetic field. condition. [D] Figure 1. Leakage Field at a Crack
- HTML** (4.47 KB): Research Development, Test Verification
Subject: Magnetic Field Distribution...-induced normal magnetic field component magnetic current sheet... magnetic field exists in between two conducting sheets, v Recommendation(s): The distribution of flaw-induced normal magnetic field compo
- PDF** (119.17 KB): to-0020.pdf
magnetic fields. After energizing, field magnitude... 17 23 Max. magnetic fields spe magnetic field emissions above... susceptibility #40792--Static magnetic field emis #40816--Not to demagnetize... magnetic fields tests. EMC TEST Number of Waiver compensation
- HTML** (15.01 KB, BEST PRACTICE): Aircraft, Communication Systems, Computers, Environment, Flight Operations, Flight Launch Vehicle, Parts Materials Processes, Payloads, Spacecraft, Test Article, Test Antenna terminals X X R - - RE Magnetic fields (STS payloads) X - - Sb RE AC mag - Rb RS Orbiter unintentional E-field X - - Rb RS Magnetic-field susceptibility X X R (compatibility) X X Rb,R... to exposure to the high-level emissions from

The right side of the browser window shows a sidebar with the text "SELECT A SEARCH RESULT TO VIEW IT HERE." and a background image of the Eagle Nebula. The footer of the page includes the Owl Insight logo and the text: "COMPANY CONFIDENTIAL / PROPRIETARY PILOT PROJECT DEVELOPED BY OWL INSIGHT LLC UNDER JPL SUBCONTRACT".



Plans for the 2nd Phase



Office of the Chief Engineer

- The proof-of-concept pilot system now online allows users to search a set of 1700 files that were provided by JPL to the contractor, processed, and indexed on a server. Presently, administrators or users cannot add more files.
- A follow-on contract is planned for a Media Search[®] system that is scalable (i.e., additional servers and document collections can be added to the system) and that allows administrators to select and process additional files or collections so they will be available for subsequent searches.
- The initial application for the delivered Media Search[®] system may be the NASA Engineering Network (NEN), a NASA repository under development by JPL, which presently contains 1.3 million files. The system will be expandable; for example, it could be applied to indexing and quickly retrieving a key discussion from among hundreds of hours of videotaped spacecraft design meetings.
- The planned system will provide a tool and procedures for managing file ingestion and processing, allow users to perform a search-by-meaning (concept search) where they don't know the exact words used in a document or multimedia file, and incorporate the standardized NASA taxonomies.