

Today's Moderators

- 1. Steve Pederson CEO & Chairman
- 2. Will Bushee *V.P. of Development*
- 3. Troy Mentzer Chief Navigator...trolling the deepest and darkest the Internet has to offer...finding relevant content for the great unwashed.

stevep@harvestIP.com / +1 (612) 840-6888 willb@harvestIP.com / +1 (605) 212-6396 troym@harvestIP.com / +1 (605) 261-0450

Agenda

- 1. About HarvestIP
- 2. What is the problem?
- 3. What is the Deep Web?
- 4. What Validity Search & Google Misses
- 5. Ten Things Missing from Search
- 6. How HarvestIP Harvests the Deep Web
- 7. Deep Web Report Examples
- 8. Q&A





Harvesting <u>Deep Data</u> from the Public Web, whether it is "known", "unknown" or "hidden".

Providing normalized, **Relevant Content** for analysis in the Legal Community.



The Company

HarvestIP, LLC is the legal industry's leader in harvesting high quality content from and the conventional Surface Web (Google) and the more authoritative, but largely inaccessible **Deep Web** sources ("unknown and hidden" data.)

HarvestIP was formed to serve the legal community building on the Deep Web Federated Harvesting technologies developed over the past 10 years by BrightPlanet Corporation.

Focusing on the Legal Community

The IP research market is growing rapidly.

Finding unstructured, <u>Non-Patent-Literature</u> (NPL) on the web is fast becoming THE key essential "missing link" for IP researchers.

HarvestIP is in a position to quickly become the leading "real-time" research resource for litigators, analysts, and researchers from the Public Web.



A Proven History

With over 10 years of Deep Web extraction expertise, and four years of direct experience working with <u>U.S. Intelligence Agencies</u>, HarvestIP's parent technology company, BrightPlanet Corporation, has achieved a strong reputation and is acknowledged as the resource for Deep Web Harvesting in the Intelligence Community (IC) on behalf of the <u>U.S. Government's War on Terror</u>.

HarvestIP has been charged with providing these same Deep Data Harvesting resources to the Legal Community - successfully integrating with "best of breed" analytic and visualization tools to provide highly relevant content for the legal community from the Deep Web.



Three Problems

- 1. Data & Documents, both (structured and unstructured) on the Open Source Public Web and on private/proprietary Databases are growing exponentially...at the "scale of the Internet".
- 2. Current search technologies (like Google) simply do not have the scope or capacity to find all "unknown and hidden" <u>Deep</u>
 <u>Data</u> on the web.
- 3. Current search technologies return popular links <u>they do not</u> <u>provide content,</u> nor can they federate and normalize it into a form that is **qualified, relevant, and useable** to the projects at hand.



The Solution

To find data that is "known", "unknown", "protected" or "hidden", we must <u>first</u> rely on a technology that does not simply crawl data, but can Harvest, Federate, Normalize and Qualify data.

HarvestIP harvests and delivers fully qualified, relevant Deep Data to analysts and analytic tools from the Open Source Public Web, Proprietary and Private sources.



Surface Web vs. Deep Web Surface Data vs. Deep Data Search-Crawl-Mine vs. Harvest

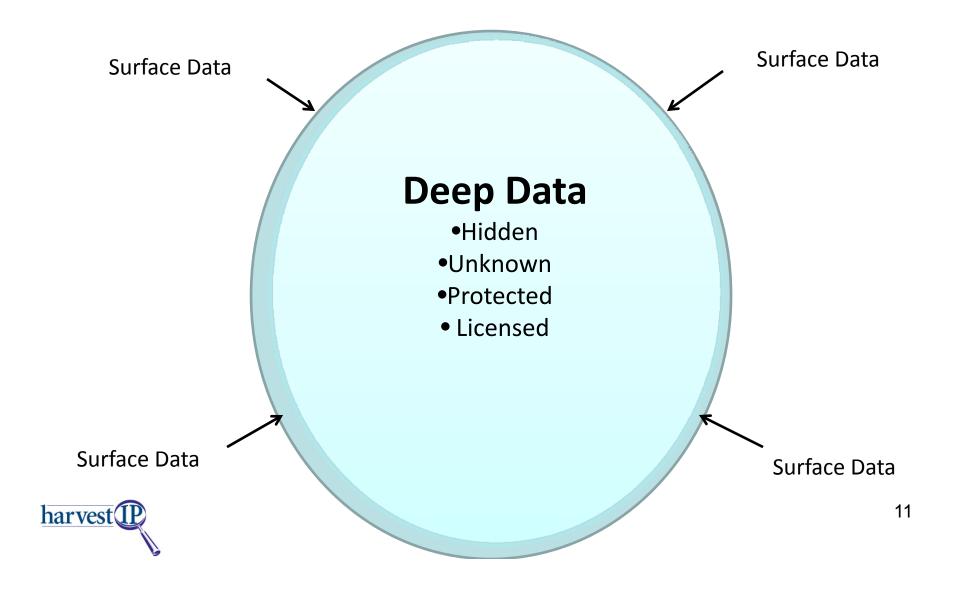
Exploring a Deep Web that Google Can't Grasp

New York Times, February 2009

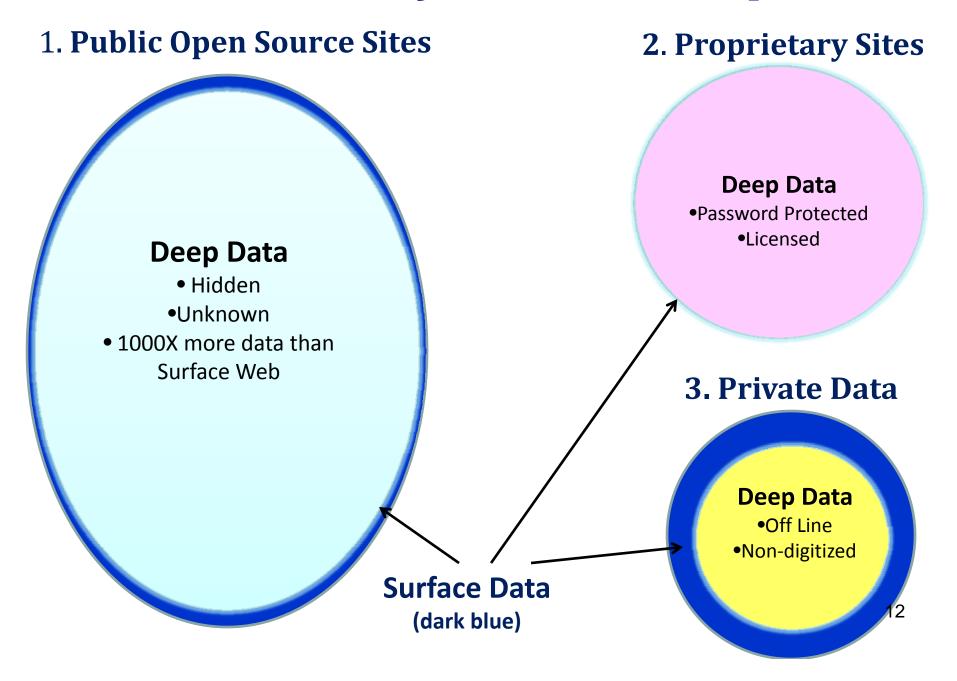
(http://www.nytimes.com/2009/02/23/technology/internet/23search.html)



Surface Data: Accessed by link traversal via Google, Yahoo!, etc. **Deep Data**: Automatically accessed by HarvestIP Technologies



3 Sources of *Surface* Data & *Deep* Data



Not Google...

Beyond "Surface Search", Crawl, & Mining

Google, Yahoo, Bing, Cuil et al: (Surface Search)

Single user searches ONLY through one search engine or index

Only one query can be issued at a time

Content may be out-of-date

Content may not exist within the index

Often too much information is returned

Often the valuable information is "hidden"



What is Validity Search Missing?

Validity Search utilizes static databases that cover patents only (such as the USPTO, Questel and Delphion). They offer a subscription to "known" data sources and have little or no access to Non-Patent Literature (unstructured Deep Data) assigned to the Web.

Deep Data on the web represents an increasingly growing repository of documents and data now generally regarded as essential to the legal diligence process in today's world.



The HarvestIP Solution

(...at the scale of the Internet)

- 1. HarvestIP <u>Harvests Documents</u> from the <u>Deep Web</u> and provides content not found through Surface Search technology.
- HarvestIP uses a patented, custom <u>configurable deep query</u> <u>technology</u> that allows the harvesting of <u>"Relevant Content"</u> documents search engines like Google cannot find.
- 3. HarvestIP directs the custom query manager to **Automatically Harvest "beyond the search box",** employing patented scripting technologies that find "unknown and hidden" Deep Data.
- 4. HarvestIP **Federates and Normalizes Content** so that it becomes relevant and purposeful to the analytic projects at hand.

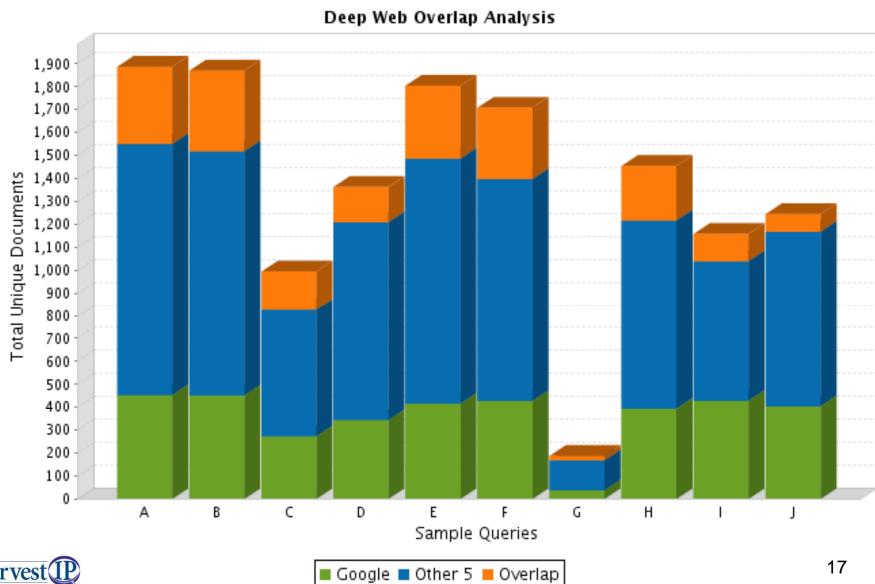


Top Ten Things...

- 1. Engines have 30B to 1T indexed...just the surface
- 2. Does "The Biggest" mean "The Best"?
- 3. Popular yes, but did I find what I was looking for?
- 4. The results depend on how you ask the questions
- 5. Different Sites...Different Results...Different Rules
- 6. Now we see it... Now we don't
- 7. The trouble finding a good date
- 8. No such thing as a "Simple Search"
- 9. Results 1-10 of about 15,000,000
- 10. How do you know what you're missing?



What Surface Search is Missing





Surface Search (Google) vs. Deep Web Harvest (HarvestIP)

An example...



Google (Surface Web) vs. Deep Web

Case Study: Searching for "Quantum Computing"

- A <u>Search</u> of Google with the query "quantum computing" returns "about 648,000" but the actual documents available are 675.
- A Deep Web <u>Harvest</u> of the Dr. Dobb's site with the identical query returns 42 documents, with all 42 available.
- A direct comparison of the two result sets showed no overlap from Google's result set to Dr. Dobb's result set.
- **Conclusion:** A researcher would not have found any of the documents contained within the Dr. Dobb's database by doing only a Google search. ("What other content was not available in the Google results for this query?")



Google (Surface Web) vs. Deep Web



FREE Subscription to Dr. Dobl Digest: Same Great Content, Digital Edition

Search

Search Result ~

Found 42 items for

Results sorted by:

Results 1 - 10

Quantum Compu

Researchers set a ne individual electron

November 24, 2008

Probe Could Aid

Spectroscopy, with a September 04, 2008

A Conversation

Erik Demaine is this research interests ra searches, to the com

September 02, 2008

A Conversation v

Christos Papadimitrio Excellence. In this ex from the economics of novelist."

Search

Search Results

Found 42 items for "quantum computing"

Results sorted by:

Date

Relevance

RE-SORT

ABOUT US | CONTACT | ADVERTISE | SUBSCRIBE | SOURCE CODE | NEWSLETTERS | RESOURCE

FREE Subscription to Dr. Dobb's

Results 41 - 42

SEARCH TIPS

For ANY words separate the terms with OR

For ONLY these words separate the terms with AND

For **EXACT PHRASES** put "Quotes Around The Phrase"

Search Help

Programming Paradigms

Michael catches up on breaking news about the 400-year old Ylbvi-Nlliv algorithm, then turns his eye to Steve Jobs efforts to pull Apple back to the top.

April 01, 1998

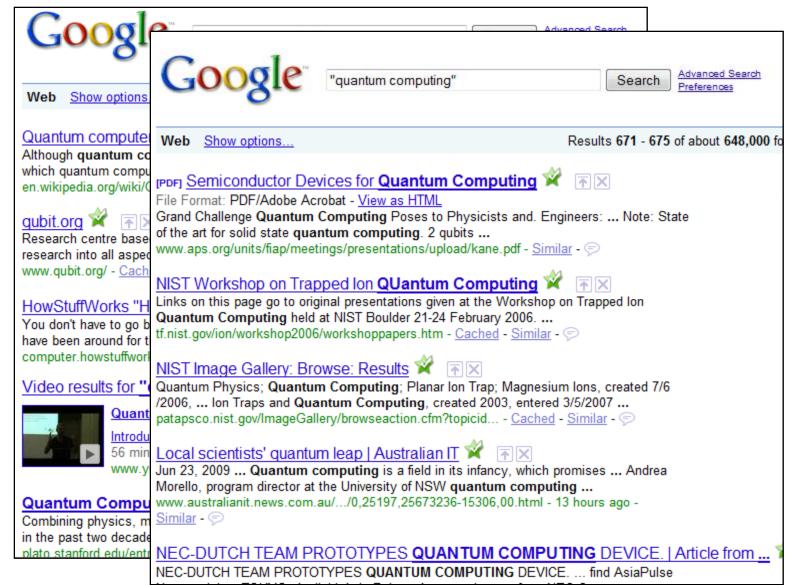
Books and Covers

What's the state of computer science today? Michael finds out, as he examines a book of the same name.

October 01, 1996

12345

Google (Surface Web) vs. Deep Web



Google "Search" vs. Deep Web Report

Example: Swartz GUI Patent Group - Prior Art Search

Google Search

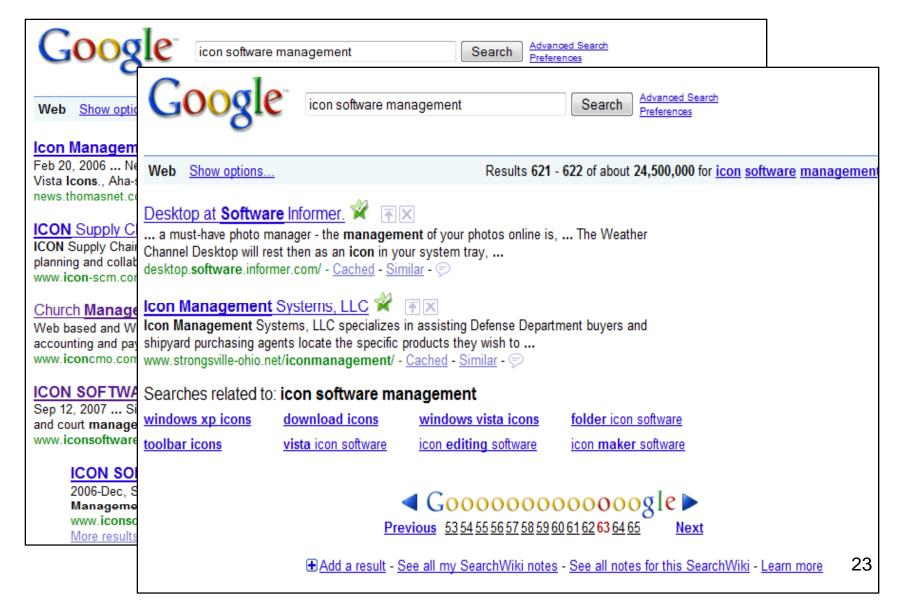
- Search only one source with one query
- Searching for "icon software management" returns "about 24,500,000 hits"
- Only 62w can be viewed

HarvestIP Deep Web Report

- Harvest from many sources
- Searching for "icon AND software AND management" to 65+ Deep
 Web sources returns 1,354 qualified results
- All are viewable and subset searchable
- Further filtering to 127 highly qualified results



Google "Search" vs. Deep Web Report



Google "Search" vs. Deep Web Report

Harvest ID: icon AND software AND management-O20090609_150648 Query: icon AND software AND management (All of these words) [Standard]								
Sources: 66 and Common								
Filters: 32 Incl Filter Te Run/Modified: Jun 10, 2009 10 Stats: 1354 docs [197 Notes: (1998 OR 1999 1-15 16-30 31-45 46-60 61-	Report Name: Icon Software Management Query: icon AND software AND management (All of these words) [Standard] Filters: 32 Incl Filter Terms; 8 Excl Filter Terms; With Domain Filters Created/Modified: Jun 24, 2009 11:25:41 AM CDT Stats: 127 accepted; 0 projects; 1 document sets Notes:							
	1-15 16-30 31-45 46-60 61-75 >> >>> >>>							
Select All (0 Selected) Language 1. International Conference Summary: Practical Software Project Management Management: SPM-ICON 2004 A Conference on Practical Software Allconferences.com > Computer International Conference on Practical May shashi@qsitglobal.ccategory 3 Computers: Object_conttp://www.allconferences.com Relevance: Size: 19k Ha Top of Page	1. PSM Insight: The Army-DoD Tool to Implement IssueDriven Software Measurement Don Scott Lucero U.S. Summary: Implement Issuedriven Software Measurement Don Scott Lucero U.s. Army Software Metrics Office The Army Software Metrics the software process. Software measures are tailored to reflect the existing project management and software development the PSM Guide, Practical Software Measurement: A Foundation for Objective Project Management. The guide explains support an integrated DOD software measurement strategy. PSM Insight provides a Pc-based management capability to implement software measures to unique project issues and allows using data already available from an existing software for a software measurement project and can manage data according to the attributes and software components,							
2. ICON plc Selects Tra								
Corrective Preventive Action Man Management Software Home S	2. Quick Tips to Using I-DEAS I-DEAS™ Tutorials: Fundamental Skills Learn about: •••••• Summary: the top of the icon panel when the software is running. For example, click on the Polylines icon with the left mouse button. Things to notice Notice that the Polylines icon is selecting the Lines icon from the stack. When you do, the Lines icon comes to the Terminate the Lines icon by pressing the middle mouse button, to deactivate icon Things to notice Part, Pull down the icon stack and select the Name Parts icon.							
Notes: (1998 OR 1999 1-15 16-30 31-45 46-60 61- Go to Result: Select All (0 Selected) Language Project Management. Management: SPM-ICON 2004 A Conference on Practical Software Allconferences.com > Computer International Conference on Practical Software Allconferences.com > Computer International Conference on Practical Software International Conference Inter	Created/Modified: Jun 24, 2009 11:25:41 AM CDT Stats: 127 accepted; 0 projects; 1 document sets Notes: 1-15 16-30 31-45 46-60 61-75 >> >>> >>> >>> Go to Result: 1. PSM Insight: The Army-DoD Tool to Implement IssueDriven Software Measurement Don Scott Luce U.S. Summary: Implement Issuedriven Software Measurement Don Scott Lucero U.s. Army Software Metrics. Office The Army Software Metrics the software process. Software measures are tailored to reflect the existing project management and software development the PSM Guide, Practical Software Measuren A Foundation for Objective Project Management. The guide explains support an integrated DOD software measurement strategy. PSM Insight provides a Pc-based management capability to implement software measures to unique project issues and allows using data already available from an existing software software measurement project and can manage data according to the attributes and software components, http://www.stsc.hill.af.mil/crosstalk/1999/06/lucero.pdf Relevance: Type: PDF Size: 127k Harvested: Jun 10, 2009 10:16:46 PM CDT Original Text Top of Page 2. Quick Tips to Using I-DEAS I-DEAS™ Tutorials: Fundamental Skills Learn about:							

Examples of a HarvestIP Prior Art Search

Deep Web Harvest Reports &

HarvestIP Searchable Portals



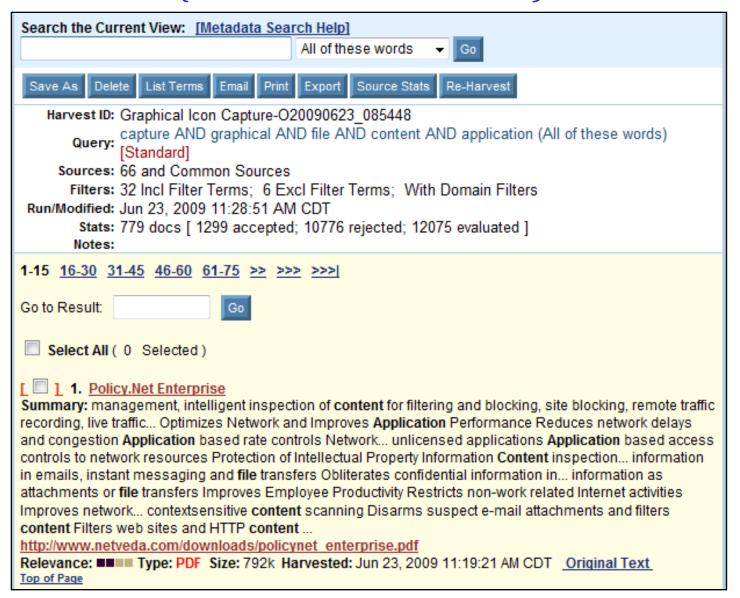
Sunlight Research Portfolio Claim Map

Interactive Map for F&G Research's Mouse Patents																		
				Me	ouse and	method	for	Comp	outer	Mo	ouse and	method	for	Mouse	driver	Comm	ands fun	nctions
				concur	rent curs	or positi	on and	keyboa	rd with	concur	rent curs	or positi	ion and	arrang	ement	invoked	from mo	vement
Similar				scrolling control				dial for e	entering	scrolling control				for providing		of a control input device		
₹								repetiti	ve data					adva	nced			
SUNLIGHT RESEARCH™								and con	nmands					scro	lling			
SUNLIGHT RESEARCH™								1 1						capabilities to				
The Source for Patent Analysis														a conventional				
,													mouse					
Title																		
File#																		
Patent#				5,313,229			5,438	3,331	5,374,942				5,633,657		5,745,719			
Assignee Name																		
Filing Date					Feb 05	, 1993		Jun 11	, 1993		Aug 12	, 1993		Oct 11	, 1994	Ja	n 19, 199	95
Total Claims				17			9		18			19		23				
Claim#(Independent)				1	9	12	17	1	9	1	6	11	16	1	11	1	8	16
Claim Status (1=don't need, 2=want, 3=need)																		
				Possibly Need		Possibly Need		Possibly Need			Possibly Need		Possibly Need					
Scope Concept (Appearance, Asc.)	Ranking (1-3)	Total		6	6	8	8	3	3	7	10	10	11	8	8	4	4	3
controlling a scrolling movement of information on a display	3	10	67%															
controlling movement of a cursor or position indicator on a display	3	13	87%															
generating x-y position signals or incremental movement information	3	10	67%															
providing a displaceable or moveable body and spring for generating control																		
signals	3	3	20%															
providing a supplementary control signal for scrolling information on a																		
display	3	8	53%															
varying control signals based on a displacement amount and direction from	_	_	2007															
equilibrium	3	3	20%															
analyzing a trail or movement of a cursor at periodic time intervals	3	7	47%															
detecting a dominant cursor direction	3	5	33%															
generating commands for scrolling information on a display in a scrolling direction along a scrolling axis	3	5	33%															
setting a scrolling axis		4																
Security a scroning axis	3	4	27%															
generating an output signal according to an angular rotation speed of a dial	3	2	13%															
modifying a sequence of repeated commands by activation of a key when a		-	1370															
	_	2	13%															
dial is rotated	3																	
dial is rotated producing a sequence of repeated commands with a rotatable dial	2	2	1204															

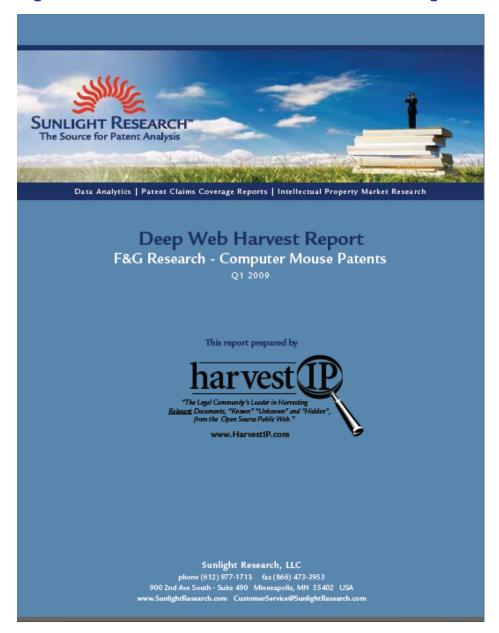


Validity Search for Portfolio Claim Map

(Initial Harvest Results)



Deep Web Harvest Report



Deep Web Harvest Report

Deep Web Harvest Report F&G Research - Computer Mouse Patents



Quarter 1 - March 2009

Document Summaries

F&G Research - Computer Mouse Patents Document Summaries

Set forth below for each document obtained in the Deep Web Harvest are Document Summaries, which include the document Title, an automatically generated document description, last harvested date, source that the document is located, and a URL link. Each document has been hand-vetted for potential relevancy to this portfolio. All documents referenced in this Deep Web Harvest Report are included in the Harvested Documents folder as a PDF file.

HIP F&G0001

ISSUES FOR PROXIMATE USER INTERFACES Ken Pier Information Sciences and Technology Laboratory Xerox

Summary: display cursor. The user could then invoke an operation from that menu by releasing the mouse ... cursor in response to a particular mouse button press, or a mouse button press without cursor ... equally close to the cursor when the menu appears, and very little cursor motion is needed... the mouse, so that existing applications for the Windows environment may simply substitute pen for mouse. ... the location of the mouse cursor to provide input focus and to direct user actions such... cursor lies. When a user moves a mouse cursor and then lets go of a mouse ...

http://www.cs.berkeley.edu/~landay/research/publications/proximate.pdf

Harvested: Tue Jan 27 11:21:25 CST 2009

Source: UC Berkeley EECU http://www.cs.berkeley.edu/

HIP F&G0002

A Taxonomy of See-Through Tools Eric A. Bier, Maureen C. Stone, Ken Fishkin, William Buxton†, Thoma.

Summary: other hand controls a mouse cursor. The user clicks through a tool onto application objects, simultaneously... by positioning a cursor using a mouse in the right hand. Together, the cursor, toolglass sheet,... into place with the mouse cursor and then apply it. These tools have many other potential... s gaze and the mouse cursor to remain in the work area. They can be

HarvestIP Searchable Portal

Search: All of	these words Go Advanced Search							
All Documents All Documents Within Scope Concept Topic								
- Analyzing a Trail or Movement of a Cursor - Calculating a Cursor-Heading Time Variation - Controlling a Scrolling Movement of Info - Controlling Movement of a Cursor - Detecting a Dominant Cursor Direction - Determining Similarity of a Cursor Trail - Executing a Predetermined Command - Generating an Output Signal According to Dial - Generating Commands for Scrolling Information - Generating Scrolling Commands According - Generating X-Y Position Signals	- Modifying a Sequence of Repeated Commands - Producing a Sequence of Repeated Commands - Producing a Time-Varying Output Signal - Providing a Displaceable or Moveable Body - Providing a Supplementary Control Signal - Providing Synchronizing Signals with a Clock - Providing Tactile Feedback About Position - Setting a Scrolling Axis - Setting or Interpreting a Scrolling Scale - Storing a Cursor's X-Y Incremental Movement - Varying Control Signals Based on Displacement							
Scope Concept Node Results: 8896								
1-15 <u>16-30</u> <u>31-45</u> <u>46-60</u> <u>61-75</u> >> >>> >>>								
Go to Result:								
1. Optimal Compensation for Changes in Task-Relevant Movement Variability Trommershäuser et al. 25 (31): 7169 Journal of Neuroscience Summary: in Task-relevant Movement Variability Trommersh user et al.: Journal of Neuroscience 25 31 7169 Optimal Compensation for Changes in Task-relevant Movement Variability Trommersh user et al. 25 31: 7169 Journal of Neuroscience Www.jneurosci.org - The Journal of Neuroscience Monoclonal Antibodies for Neuroscience Research QUICK SEARCH: advanced Author: Keyword s: Year: Vol: Page: - HOME SEARCH ARCHIVE SUBSCRIBE CONTACT HELP The Journal of Neuroscience, August 3, 2005, 25 31:7169-7178; doi:10.1523/jneurosci.1906-05.2005 This Article Right arrow Abstract Freely available Right arrow Full Text PDF Right arrow Supplemental data Right arrow Submit an eletter Right arrow Alert me when this article is								

Harvesting "Unknown and Hidden" Relevant, Current Content for:

- Validity Search
- Patentability
- Portfolio Management
- Litigation Support
- Claims Coverage
- Portfolio Management
- Competitive Intelligence
- Licensing
- Freedom to Operate
- ...and much more.





Deep Web Harvest Reports can also be found at:



http://sunlightresearch.com

Q & A



Thanks!

Steve Pederson

stevep@harvestIP.com / +1 (612) 840-6888

Will Bushee

willb@harvestIP.com / +1 (605) 212-6396

Troy Mentzer

troym@harvestIP.com / +1 (605) 261-0450

Nora Best

norab@harvestIP.com / +1 (815) 549-6613

