

# Newspaper Preservation

by

**H.R. Mohan**

Associate VP (Systems)

The Hindu

Chennai – 600002

[hrmohan@gmail.com](mailto:hrmohan@gmail.com)

# Newspapers - An Introduction

The newspaper is a product born out of

- Necessity
- Invention
- The middle class needs
- Democracy
- Free enterprise
- Professional standards.

# Importance of Newspaper Archives

- Newspapers may perish but not the news they contain.
- The news become history.
- The greatest part of general information today is found in Newspapers.
- To trace the history and refer, people look for the newspaper archives.

# The Collections at the Newspaper office include

- The Printed copy
- Supporting Documents: facts, tables, statistics
- Photographs
- Illustrations: maps, charts
- Clippings

# Archives & Preservation

- Hard Copy
- Microforms: film / fiche
- Digital Form
  - Full Text
  - Image Files
  - HTML / XML Pages
  - PDF Files

# Retrieval of Information

- Index
- Document Management Systems
- Full Text Retrieval Systems
- CDROM based Retrieval Systems
- Digital Asset Management Systems
- Web based Internet / Intranet

# Delivery of Information

- Conventional Photocopy
- Microfilm Reader/Printer
- CDROM
- Email
- Web
- RSS Feeds
- Mobile and Handhelds
- Online Services
- Through Content aggregators

# Status of Digitisation

- Low Priority
- Unorganised
- Missing hardcopies
- Microfilm exists but quality ?
- Non availability of Reader/Printer
- Sketchy Index
- Manage with clippings
- Last few years in digital form (Born Digital)
- Rush to digitise and store in CDROM / Local Systems
- Attempts to Web Enable
- Unclear Business Model



# Digitisation & Business Issues

- Quality of originals / hardcopy
- Size of the paper
- High cost of scanners
- Format of storage: Image, PDF, HTML/XML
- Conversion: OCR, HTML/XML
- Tagging
- Indexing old issues
- Storage: CDROM / Optical / Magnetic (online)
- Period of Access
- Deployment: Intranet / Internet / Online
- Local printing for use
- Copyrights
- Fee for use: free/subscription/pay per view -- Business Model
- Reuse

# The Hindu – An Introduction

- India's National Newspaper
- Started in 1878 as a weekly
- Became a daily in 1889
- Circulation of over 1,100,000 copies
- Over 40 lakh readers
- Published from 13 Centres as 54 Editions
- Exclusive Supplements on almost all the days
- Extensive Use of Info Tech in its activities – right from News Gathering to Archives

# The Hindu – Several Firsts

- Distribution through Aircraft
- Electronic Typesetting
- Fax Editions
- Satellite Communication
- Automated Pagination
- Internet Edition

# The Hindu – Group Publications

- The Hindu Business Line - Business Daily
- The Sportstar - Weekly Sports Magazine
- Frontline - Fortnightly Features Magazine
- Survey of Indian Industry - An annual
- Survey of Indian Agriculture - An annual
- Survey of the Environment - An annual
- The Hindu Index - Monthly and Cumulated Annual
- Special Publications under the series THE HINDU SPEAKS ON Libraries; IT; Management Vol 1 & 2; Education; Religious Values; Music; Scientific Facts Vol 1 & 2
- Special Supplements

# The Hindu – Archives & Info Services

- Library
  - News Indexing
  - Photo Indexing
  - Book Reviews
  - Clipping Services
  - Full Text storage & retrieval
- Feed to Online Services
- Internet Edition
- ePaper
- Digital Photo Archives
- Digital Archives of Newspaper Volumes

# The Hindu – Index .. Contd

## The present status

- The Hindu News for 1988 & 1989
- The Hindu News from 1990
- Frontline News from 1988
- The Sportstar News from 1988
- Published Photos (covering both general and sports)
- Unpublished transparencies

# The Hindu - Manual Index

National Hockey Championships - see under hockey.  
 National Hockey. L. Srinivasan, Editor, 'The Hindu' presenting L. Ranganami  
 Memorial Trophy to President, Madras Hockey Association - May 30-8.

Nehru - members of mission of the I.C. of Free Trade Union  
 met at Jalki, Aug 11-10 bed.

" photographed with Sir Owen Dixon & Mr. Diagonal Bicklan  
 at New Delhi on July 20, July 22-6.

" opening the Patiencekol Bridge in 1908, Aug 23-12.

" meets members of the American Young Men's Association,  
 Aug 28-12.

- signing treaty of friendship & alliance - Aug 15-16

- opening Panama National Chemical Laboratory - Aug 15-16

- attended Colombo Commonwealth Quinquennial conference - Aug 15-16

- with Prime Minister of Nepal - Aug 15-16

- signing universal pact on minority problem - Madras - Aug 15-16

- addressing at Jalki - Aug 15-16 b

- with Governor of Singapore - Aug 15-16 c

- being introduced to Burmese Ministers - Aug 15-16 d

- with Sir Owen Dixon & Sir J. H. K. Khan - Aug 15-16 a

- addressing in Delhi - Red Fort - Aug 17-16

- inspecting a guard of honour - Red Fort - Aug 17-16

- addressing from the ramparts - Red Fort - Aug 17-16

- & other minutes at a function given by the Press Association,  
 New Delhi - Aug 27-16.

see page 309

# The Hindu – Printed Index

## THE HINDU INDEX

69

**Information Services & Technology (see also Computers / Libraries & Librarians)**

**Awards**  
**Stockholm Challenge Award**  
 Gyandoot Project of Dhar Dt. in MP bags award in Public Service and Democracy category, JUN.07,9a  
**Webby Awards**  
 some observations, JUN.08,25d  
**WebWiz Awards**  
 instituted by Brandwidth, award ceremony in Mumbai on Jul. 1, JUN.17,18a

**Bills & Acts**  
**Convergence Bill**  
 'super-regulatory body' planned for telecom, internet and broadcasting, JUN.19,10c  
**Information Technology Act**  
 skirting crucial issues: experts' views, JUN.22,25e  
**Information Technology Bill**  
 IT Bill '99 & beyond, art., JUN.08,12c; President gives assent, Act to be enforced in stages, 20,10d

**Careers**  
 IT's a Small World After All!, art., pix., JUN.28,29a; Career Clues, 28,29g

**Communications Networks**  
 Growing realism about convergence, art., JUN.01,25e

**Community Internet Project**  
**Tamil Nadu**  
 WorldTel chairman, Sam Pitroda and community internet scheme, ref. in Talk of the Town, pix., JUN.02,2d; WorldTel-Reliance-ELCOT project: cable-laying work started in Chennai and Madurai, 29,1b; cable laying work: companies have to pay Govt. annual 'track rents' to local bodies, stat., 29,1d

**Conferences**  
 on 'IT and Development', in New Delhi; Union IT Min. Mahajan speaks, JUN.22,19b

**Crime & Criminals**  
**New Delhi**  
 cyber-theft charges, bail plea of computer engineer Mukesh Gupta rejected, JUN.02,14d

**Exports**  
 export panel to promote IT enabled services, JUN.08,18c

**Hyderabad IT Engineering Consultancy City:**  
 see 'HITEC City' above

**Indian Institute Of Information Technology**  
 TN Chief Minister's plea to Centre to set up in Chennai on the lines of Gwalior, JUN.12,1b

**Information Kiosks**  
**Chennai**  
 Chennai Corporation to open information kiosks, a report, JUN.13,3b

**Information Technology**  
 Bill Gates says in the next 10 years IT would bring computers with much more powerful memory, JUN.14,16a; use IT for human development: UN official, 21,19b; use IT to tackle social challenges - Mahajan, 22,19b

**Tamil Nadu**  
 international data links throughout the State planned: CM, JUN.12,1b; TN is well poised to emerge as the largest IT

**Information Services & Technology**

**Information Technology**  
**Tamil Nadu (Contd.)**  
 State in India: CM quotes Harvard University report, 12,1b; Union Minister Pramod Mahajan compliments initiatives taken by TN Govt., 12,1d

**Information Technology Industry**  
 PC segment drives IT growth: IDC survey, JUN.17,19b

**Information Technology Parks**  
**Chennai**  
 TIDEL Park in Taramani inauguration and PM's visit, ref. in Talk of the Town, pix., JUN.14,2f; Opposition Councillors allege foul play in road work contract in TIDEL Park, 19,3b; TIDEL Park: the facts speak for themselves, advt., pix., 25,7a; TIDEL Park's inauguration on July 4, advt., pix., 30,7a

**Karaikudi**  
 CECRI plans to set up CECRI-IT Park, JUN.03,3b

**Tamil Nadu**  
 ELCOT invites business houses/entrepreneurs to set up IT Parks, JUN.11,20g

**Information Technology Policy**  
 Min. Pramod Mahajan calls meeting of various States on Jul.15 to work out policy by Sep., JUN.26,1d

**Internet**  
 VSNL announces 50 pc cut in rates and free access at nights as part of 'monsoon package', JUN.01,1a; VSNL 'monsoon package' for customers, rates cut by 50 p.c., 01,10e; Virtual private network, art., 01,25a; high-performance fibre optic detector: OptoLynx, 01,29g; When software is the service: Application Service Providers, art., pix., 01,30a; Wireless net may require 'smart antennas', art., 01,31e; new software from Computer Associates to block e-mail worms, 02,19c; Opensoft offers Internet sans PC, 08,18f; Web escorts: Naviscope, art., pix., 08,25a; MTNL announces 30 pc cut in tariffs, 10,1e; A history of tomorrow's letters: message converted into codes, art., cari., 10,2a; The sweet XML of success: language of web pages, art., 15,30f; Cyberwave plans to offer 'Internet through cable service', 16,22a; Taiwanese businessman devises way to increase speed of transmission from megabytes to gigabytes, 18,1a; more bytes in less time: Taiwanese businessman Ding Yuan Yang's venture, 18,11e; IT returns through net soon: CBDT Chairman Balasubramanian, 20,5g; An address book for the road, art., 22,25a; the other side of Net: porno-pictures being viewed by children, Itr., 29,12f

**Chennai**  
 More on the job: Jobkhoz website launched by WeaveIT of USA, art., JUN.03,2e; Easier than e-mail: Bharatmail website launched, art., 08,2e; Watchful eyes in a cyber world, art., pix., 08,2f; internet subsidy on bank loans for ex-servicemen increased, 10,3a; DoT to introduce inter-node switches to alert Internet users for incoming



# The Hindu – Photo Archives – Query & Result

MS-DOS Prompt - 10

Auto

ISYS 3.1b Database D:\H5\ISYSPH\ Enter Search

Command Menu-A clinton  
Query- Saved-Query List  
Options  
Exit

87 Hits in 64 Documents

D:\H5\ISYSPH\PHOTO.DBF	1063	(1 Hits)
D:\H5\ISYSPH\PHOTO.DBF	1024	(1 Hits)
D:\H5\ISYSPH\PHOTO.DBF	1021	(2 Hits)
D:\H5\ISYSPH\PHOTO.DBF	1002	(1 Hits)
D:\H5\ISYSPH\PHOTO.DBF	990	(2 Hits)
D:\H5\ISYSPH\PHOTO.DBF	978	(1 Hits)
D:\H5\ISYSPH\PHOTO.DBF	975	(3 Hits)
D:\H5\ISYSPH\PHOTO.DBF	968	(1 Hits)
D:\H5\ISYSPH\PHOTO.DBF	959	(2 Hits)
D:\H5\ISYSPH\PHOTO.DBF	952	(1 Hits)
D:\H5\ISYSPH\PHOTO.DBF	927	(1 Hits)

Hit List

CLINTON 87  
Final Result..... 87

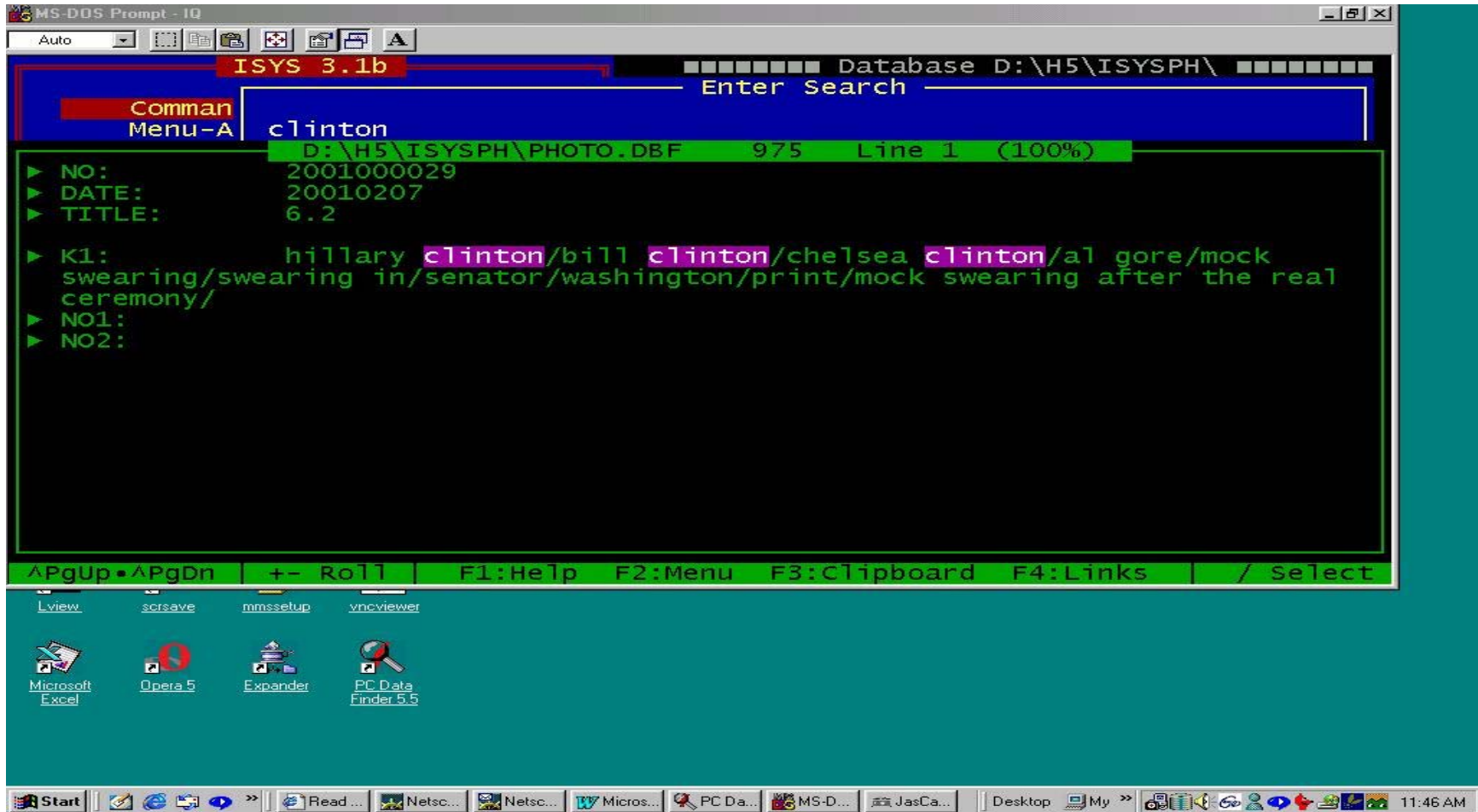
|| Move Ptr | ← select | F2: Advanced | → Hit List | ESC: Exit

Lview scrsave mmssetup vncviewer

Microsoft Excel Opera 5 Expander PC Data Finder 5.5

Start | Read... | Netsc... | Netsc... | Micros... | PC Da... | JasCa... | MS-D... | Desktop | My >> | 11:44 AM

# The Hindu – Photo Archives – Conventional - Photo Details





# The Hindu – Photo Archives – NICA - DAM System - Images

(User: acm) Image Archive Browser - Gallery : Today Pictures

File Edit Browser View Objects Utility

Search Conditions

Creation Date

Arrival Date

Image ID

Product

Slug

Looking for:

Publication Date

Sort Field

Desc Asc Incoming Date

Query log

Show >>>





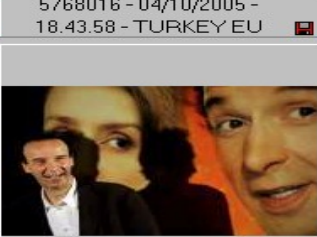


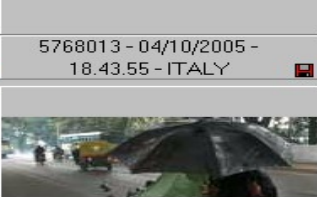

Browse Search

Images 3

Focus on: IconBoard

Selections: 0 1-8/2629

Start WS\_FTP Pro 10.10... (User: acm) Ima... 6:45 PM

 5768017 - 04/10/2005 - 18.44.49 - IBM EMPLOYEES	 5768016 - 04/10/2005 - 18.43.58 - TURKEY EU	 5768015 - 04/10/2005 - 18.43.58 - ITALY SOCCER
 5768014 - 04/10/2005 - 18.43.57 - IND0491A.JPG	 5768013 - 04/10/2005 - 18.43.55 - ITALY	 5768012 - 04/10/2005 - 18.43.55 - ITALY
		

**IMGS Id: 5768017**  
**Caption: Employees of IBM India were evacuated from Prestige Towers on Residency Road after fire broke out due to short circuit**

# The Hindu – Photo Archives – NICA - DAM System - Graphics

The screenshot displays the NICA Graphic Archive Browser interface. The window title is "(User: acm) Graphic Archive Browser - Gallery : Inbox". The interface includes a menu bar (File, Edit, Browser, View, Objects, Utility), a toolbar with navigation icons, and a left-hand navigation pane. The navigation pane shows a tree structure under "Nica" with folders like "Galleries", "Archive [GRPH]", "Others [GRPH]", "Workflow [GRPH]", "CCI", "Inbox", "Migrating to Permanent", "Proposed for Archiving [PA]", "Proposed for Publishing [P]", "Source", "Saved Queries", "Collections", "Applications", "Export Directories", and "Printers". The main area shows a grid of graphic thumbnails. The top-left thumbnail is highlighted with a red border. Below the grid, a byline section displays: "Byline: GRPH ID: 5767292 Arrival Date: 04/10/2005 File Name:". At the bottom, a status bar shows "Graphics 3", "Focus on: IconBoard", "Selections: 0", and "1-6/1999". The Windows taskbar at the very bottom shows the Start button, several icons, and the system tray with the time "6:47 PM".

**(User: acm) Graphic Archive Browser - Gallery : Inbox**

File Edit Browser View Objects Utility

Nica

- Trash
- Galleries
  - Archive [GRPH]
  - Others [GRPH]
  - Workflow [GRPH]
    - "3 Days To Go"
    - All Temp Graphics
    - Inbox
    - Migrating to Permanent
    - Proposed for Archiving [PA]
    - Proposed for Publishing [P]
    - Source
- Saved Queries
- Collections
- Applications
- Export Directories
- Printers

Refresh

Collection Informations

User:

Num. Obj:

Collection Permissions

Browse Search

Graphics 3

Focus on: IconBoard

Selections: 0 1-6/1999

Start WS\_FTP Pro 10.10... Clipboard01 - lrfan... (User: acm) Gra... 6:47 PM

**Byline:**  
**GRPH ID: 5767292**  
**Arrival Date: 04/10/2005 File Name:**

# The Hindu – Photo Archives – NICA - DAM System - Pages

The screenshot displays the 'Page Archive Browser' application window. The title bar reads '(User: acm) Page Archive Browser - Gallery : Last 3 months'. The interface includes a menu bar (File, Edit, Browser, View, Objects, Utility), a toolbar with various icons, and a left-hand navigation pane. The navigation pane shows a tree structure under 'Nica', including 'Trash', 'Galleries', 'Archive [PAGE]', 'Last 3 months', 'Others [PAGE]', 'Workflow [PAGE]', 'Saved Queries', 'Collections', 'Applications', 'Export Directories', and 'Printers'. The main area is a grid of newspaper page thumbnails. The selected page, highlighted with a red border, is identified by the following information:

**PAGE ID: 5766135**  
**Publication Date: 04/10/2005**  
**News Paper: FR**

At the bottom of the window, there is a status bar showing 'Pages: 3', 'Focus on: IconBoard', 'Selections: 0', and '4-9/33164'. The Windows taskbar at the very bottom shows the Start button, several application icons, and the system tray with the time '6:48 PM'.

# The Hindu – Photo Archives – NICA - DAM System - Text

The screenshot displays the NICA - DAM System interface. On the left is a tree view of the archive structure, including folders like 'Galleries', 'Others', 'Workflow', and 'CCI'. The main area shows a list of objects with columns for 'Arrival D...', 'Arrival ...', 'Object Name', and 'Memo'. Three objects are visible, each with a red highlight on its title:

- 16/07/2004 - 02:43:12 - **<body><digest\_head>Intelligence {lsquo}{lsquo}flawed"<252></digest\_head> <dateline>London, July 14.</dateline> <ragged\_right><semi\_bold>An official**
- 16/07/2004 - 02:43:09 - **<head>Political nexus in stamp paper scam under scrutiny</head><body><byline>By S. Vijay Kumar<252></byline><dateline>MADURAI, JULY 14.**
- 16/07/2004 - 02:43:02 - **<strap line>Centre to file application in**

An inset window titled '3125159 - 16/07/2004' shows the full text of the first object:

*Intelligence "flawed"*

*Intelligence "flawed" London, July 14. An official inquiry by Lord Butler into the quality of British intelligence on Iraqi weapons of mass destruction said today some sources were "seriously flawed" or "unreliable" but found no evidence of "deliberate distortion or culpable negligence." [mdash] AP Details on Page 13 Armitage apologises New Delhi, July 14. The U.S. Deputy State Secretary, Richard Armitage, today apologised to George Fernandes, for strip-searching done on him during an official visit to the U.S. as the Defence Minister and said he was "horrified" at the incident. [mdash] PTI See also Page 11 Petition admitted New Delhi, July 14. The Supreme Court today issued notice to the Union Government on a Public Interest Litigation, challenging the*

The taskbar at the bottom shows the Start button, several application icons, and the system clock at 8:59 PM.

# The Hindu – Images on the Web - Home

The screenshot shows a Mozilla Firefox browser window displaying the homepage of 'The Hindu Images'. The browser's address bar shows the file path 'file:///J:/npdl/th-images.htm'. The website's navigation menu includes 'Home', 'My Profile', 'Advanced Search', 'Wishlist', 'Shopping Cart', 'Purchase History', and 'Gift Certificate'. The main content area features a large image of a pink flower with green leaves. To the left of the main image are four category buttons: 'CREATIVE', 'NEWS', 'SPORTS', and 'HISTORIC'. Below these is a 'Special Collection' dropdown menu. To the right of the main image are three utility boxes: 'Members Sign In' with fields for 'Username' and 'Password', a 'go' button, and links for 'Register' and 'Forgot Password'; 'Keyword Lookup' with a search field and a 'go' button, and a link for 'Advanced Search'; and 'Quick Purchase Help' with an 'Enter Order Reference' field and a 'go' button. At the bottom, there is a 'Recent Additions' section with links for 'Creative', 'News', 'Sports', and 'Historic'. The footer contains links for 'About us', 'Site Map', 'Help', 'License and Terms', 'Privacy Policy', 'Copyright and Disclaimer', and 'Contact Us'.



# The Hindu – Images on the Web - Historic

The Hindu images - Mozilla Firefox

File Edit View History Bookmarks Tools Help

file:///J:/npd/th-images-historic.htm

Getting Started Latest Headlines

**THE HINDU**  
**images**

Home | My Profile | Advanced Search | Wishlist | Shopping Cart | Purchase History | Gift Certificate

**Other Collections**

- Creative Collection >>
- News Collection >>
- Sports Collection >>

Get the best from our Historic collection



**Historic Exclusive**

Get it from The Hindu Images.  
*online gallery of rights managed photographs*

**Welcome to our exclusive collection of historic photographs.**

You don't have to dig far and wide to find historic photographs -- Find them here. The Hindu Images uses the famous Hindu Photo Archives and The Hindu's 125-year history to provide you with some of the rarest historic photographs. Use them in books, posters, interiors, ad campaigns... anything you can imagine.

**Members Sign In**

Username

Password  go

Not a member? Register

Forgot Password

**Browse All**

**Search this section**

Keyword or image number:  
 go

Enter the keyword that describes your image

Specify Date Range: (dd-mm-yyyy)

StartDate:

EndDate:  go

**Quick Purchase Help**

Enter Order Reference  
 go

**Recent Additions**

[Creative](#) [News](#) [Sports](#) [Historic](#)

About us | Site Map | Help | License and Terms | Privacy Policy | Copyright and Disclaimer | Contact Us

# The Hindu – Images on the Web - Tsunami

The Hindu images - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.thehinduimages.com/hindu/mySearch.do









Getting Started Latest Headlines

The Hindu images The Hindu images

**THE HINDU images** Home | My Profile | Advanced Search | Wishlist | Shopping Cart | Purchase History | Gift Certificate

Search Result: TSUNAMI 1788 images found.

Page 1 of 224 Go to Page:  go [List View](#)

 <b>4049923</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>4057928</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>4051636</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>4051630</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>
 <b>4090981</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>4051145</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>4051504</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>4096394</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>

Page 1 of 224 Go to Page:  go [List View](#)

Search Executed Time: 0 Seconds

About us | Site Map | Help | License and Terms | Privacy Policy | Copyright and Disclaimer | Contact Us

**Members Sign In**

Username

Password  go

Not a member? Register

Forgot Password

**Search the site**

Enter PhotoId / Keyword(s):  go

[Advanced Search](#)

**Quick Purchase Help**

Enter order reference  go

**Special Collection**

Choose from Collection

# The Hindu – Images on the Web – Tsunami - Chennai

The Hindu images - Mozilla Firefox

File Edit View History Bookmarks Tools Help

file:///D:/npd/th-images-tsunami-chennai.htm


Getting Started Latest Headlines

The Hindu images

**THE HINDU**  
images

Home | My Profile | Advanced Search | Wishlist | Shopping Cart | Purchase History | Gift Certificate

[Back to collection](#)



[Add to cart](#)  
[Add to wishlist](#)  
[Calculate price](#)

**Photo Id:** 4050403  
**Web Category:** NEWS  
**Category:** DIS  
**Sub Category:** EARTHQUAKE

**Keyword List**  
EARTHQUAKE, UNDERSEA EARTHQUAKE, DISASTER, TSUNAMI, PEOPLE, DAMAGE, TIDAL WAVE, FISHING BOATS, CHENNAI

**Caption:** Damaged fishing boats at Ennore, due to the tidal wave that lashed the Chennai coast on December 26, 2004. Tsunami or huge seismic sea waves that were triggered by a massive undersea earthquake off Sumatra in Indonesia, had left over 9,300 people dead and tens of thousands homeless in India, Sri Lanka and South-East Asia. Photo: K.V. Srinivasan

**Photographer :** K V Srinivasan  
**Credit:** THE HINDU  
**Creation Date:** 26-12-2004  
**City:** CHENNAI  
**State:** TAMIL NADU  
**Country:** INDIA

**Image Width:** 2127 pixels  
**Image Height:** 1420 pixels  
**File Size:** 557368 bytes

**Members Sign In**  
Username   
Password    
Not a member? [Register](#)  
[Forgot Password](#)

**Search the site**  
Enter PhotoId / Keyword(s):  
   
[Advanced Search](#)

**Quick Purchase Help**  
Enter order reference

**Special Collection**

[About us](#) | [Site Map](#) | [Help](#) | [License and Terms](#) | [Privacy Policy](#) | [Copyright and Disclaimer](#) | [Contact Us](#)

# The Hindu – Images on the Web – Actresses

The Hindu images - Mozilla Firefox

File Edit View History Bookmarks Tools Help

file:///D:/npdl/th-images-actress.htm

Getting Started Latest Headlines









The Hindu images The Hindu images

**THE HINDU**  
**images**

Home | My Profile | Advanced Search | Wishlist | Shopping Cart | Purchase History | Gift Certificate

Search Result: ACTRESS 694 images found.

Page 1 of 87 Go to Page:  go List View

 <b>3932668</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>3986235</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>735026</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>3495464</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>
 <b>3331017</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>1350184</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>1970524</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>	 <b>1666191</b> <a href="#">Add to cart</a> <a href="#">Add to WishList</a> <a href="#">Calculate Price</a> <a href="#">Show Details</a>

Page 1 of 87 Go to Page:  go List View

About us | Site Map | Help | License and Terms | Privacy Policy | Copyright and Disclaimer | Contact Us

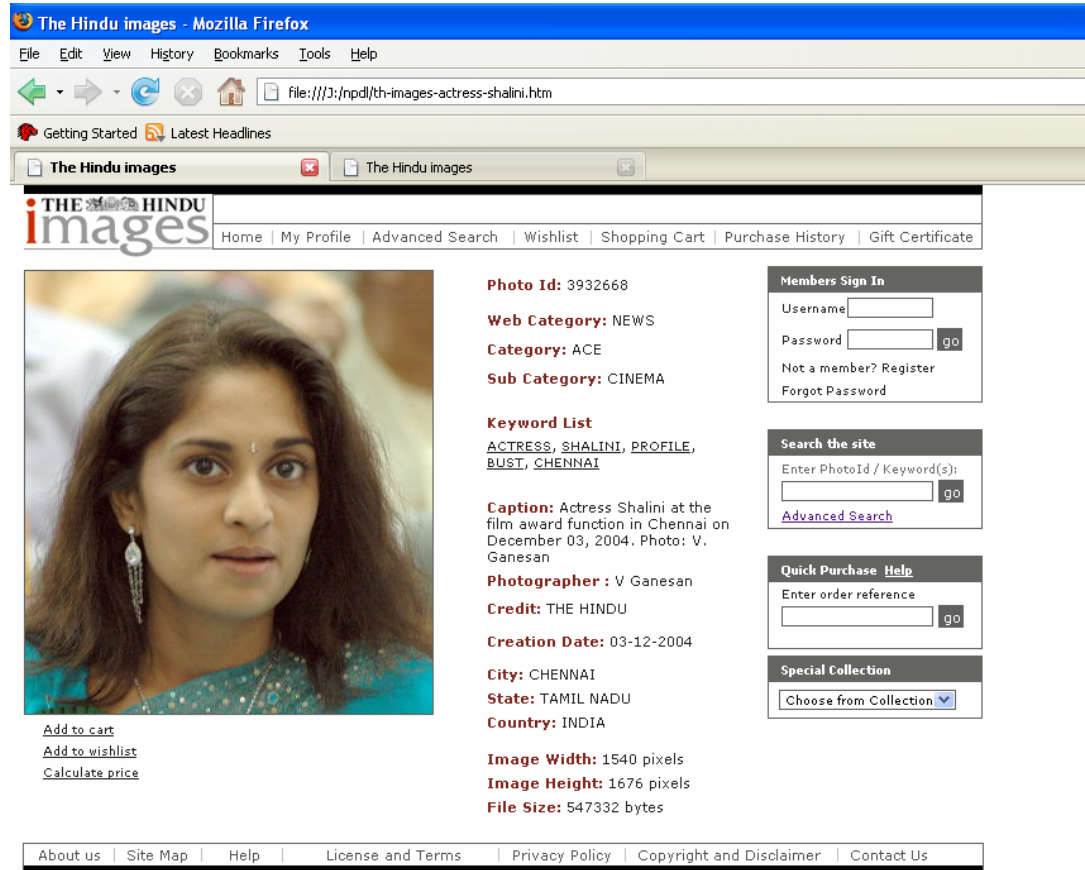
**Members Sign In**  
Username   
Password  go  
Not a member? Register  
Forgot Password

**Search the site**  
Enter PhotoId / Keyword(s):  
 go  
[Advanced Search](#)

**Quick Purchase Help**  
Enter order reference  
 go

**Special Collection**  
[Choose from Collection](#)

# The Hindu – Images on the Web – Actresses - Shalini



The screenshot shows a Mozilla Firefox browser window with the address bar displaying 'file:///3:/npd/th-images-actress-shalini.htm'. The website header includes the logo 'THE HINDU images' and a navigation menu with links for Home, My Profile, Advanced Search, Wishlist, Shopping Cart, Purchase History, and Gift Certificate.

The main content area features a large photograph of actress Shalini. To the right of the photo, the following metadata is displayed:

- Photo Id:** 3932668
- Web Category:** NEWS
- Category:** ACE
- Sub Category:** CINEMA
- Keyword List:** ACTRESS, SHALINI, PROFILE, BUST, CHENNAI
- Caption:** Actress Shalini at the film award function in Chennai on December 03, 2004. Photo: V. Ganesan
- Photographer:** V Ganesan
- Credit:** THE HINDU
- Creation Date:** 03-12-2004
- City:** CHENNAI
- State:** TAMIL NADU
- Country:** INDIA
- Image Width:** 1540 pixels
- Image Height:** 1676 pixels
- File Size:** 547332 bytes

Below the photo are three links: [Add to cart](#), [Add to wishlist](#), and [Calculate price](#).

On the right side of the page, there are several utility boxes:

- Members Sign In:** Includes fields for Username and Password, a 'go' button, and links for 'Not a member? Register' and 'Forgot Password'.
- Search the site:** Includes a search input field, a 'go' button, and a link for 'Advanced Search'.
- Quick Purchase Help:** Includes a field for 'Enter order reference' and a 'go' button.
- Special Collection:** Includes a dropdown menu labeled 'Choose from Collection'.

The footer of the page contains a navigation menu with links for About us, Site Map, Help, License and Terms, Privacy Policy, Copyright and Disclaimer, and Contact Us.

# The Hindu Archives - Preservation

- Initiative Started in 2001
- Preservation was the key requirement as the paper was losing strength and handling for reference became difficult as it was crumbling
- The manuscript Index volumes numbering 3000+ also became difficult to handle for periodic reference
- Strengthening the paper was planned
- Thin muslin cloth bonding preferred over lamination
- About 1.2 million pages were strengthened over a period of FOUR years
- It also facilitated to know the inventory of our holdings

# The Hindu Archives - Digitisation

- The preservation activity had limitations of access
- For better access & retrieval of information Digitisation was considered to be the solution
- In 2003 a working group was formed with Dy. Chief Librarian & Chief Systems Manager under the guidance of Editor & Joint Managing Director to study and initiate a project
- Considerable cost (multi crore) was projected
- Initial trials were done at CDAC, Bangalore where a pilot project for IIAP was being carried out
- CDAC was more towards book digitisation
- The newspaper digitisation involved segmenting the news and advertisements and building up databases and explicit search & retrieval facilities

# The Hindu Archives - Digitisation

- Search was initiated to locate agencies who can digitise the large size & newspapers in high volume and also use the microfilms as input wherever hardcopy was not available / in poor condition plus work on digital pdf files as well
- Out of Six agencies identified, three were dropped as they were not geared up for the full digitisation process up to retrieval interface
- Two agencies from Chennai and One agency from Hyderabad were short listed for demonstrating Proof of Concept.
- At a broad scale deliverables were defined as [POC Specs](#)
  - Full Page Image (in Tiff & jpg format)
  - Full Page PDF (image over text form)
  - Splitting Individual Stories & OCRing (pdf, jpg & XML form)
  - Splitting individual photos & advertisements (pdf & jpg form)
  - Tagging the XML stories
  - Simple retrieval system using Open Source Software



# The Hindu Archives - Digitisation

- One Agency from Hyderabad & one from Chennai were short listed
- Commercials & their similar work project experience were considered for finalising the contract.
- Pre-condition was that the originals will not be shifted from the office
- Both the agencies were very aggressive as The Hindu was a prestigious client
- Considering the ease of co-ordination, the Chennai based agency was awarded the contract to demonstrate & develop a prototype based on the first version of the specifications so that we can refine our specs
- A sample lot of about 5000 pages spanning at an interval of Five years from the inception 1890 to 2000 were used in the prototype. This gave us an idea of the newspaper layout, content organisation, other elements etc. This was very valuable in arriving at the project [specifications](#). We referred to NewsML & IPTC standards too.

# The Hindu Archives - Digitisation

- Final specs were frozen in Aug 2003 and the order was confirmed
- Workspace for 10 people and two A0 size scanners were provided for the contractor to work on the project
- Library staff co-ordinated with the issue of the hard copy newspapers for scanning
- After scanning (two pages at a time) the files were split, cleaned and stored in TIFF format
- Periodically the TIFF files were sent to the Data Centre of the contractor (outside our office) for creating the project deliverable components
- The deliverables and associated files were stored date wise and a database was created and stored on a staging server at The Hindu to facilitate the Quality Check process
- Library staff were trained by the Systems Dept on how to check the digitised Pages, News Items, OCR'd Text, Metatags, Advts and the related Links etc
- Corrections were carried out by the contractor personnel on the local staging server
- From Staging Server, the data was transferred to SAN attached to NICA

# The Digitisation Workflow

- [Generic Workflow of Newspaper Digitisation](#)

# Digitisation -- Issues

- Missing Pages
- Foldings, ink streaks, pasting with tapes
- Cutting at edges (as pages were trimmed during binding)
- Multiple editions -- Overlapping of contents
- Scanning problems in strengthened pages
- Problems with Microfilms (storage, filming patterns)
- Scanning problems in pages printed from Microfilms -- inconsistent exposure
- OCR related issues for the earlier period items
- News items with no title
- Identifying items for zoning as too many short news items – two / three lines
- Meta Tagging – lack of experience – clarity
- Quality Check – tedious and time consuming

# Digitisation -- Storage

- Up to 20 MB for the TIFF files per page and about 40 MB for the components per day (avg) for the older periods and now it is in the range of 80-100 MB because of more pages and colour printing
- Large Storage Volume anticipated – 35 TB but expected to expand to 50 TB
- Original scanned TIFF Files on CD / DVD / Tape Cartridge – to reduce cost
- Full page PDF/JPG and components on staging server for Quality Check
- Backup of components on Tape Cartridge
- Corrected files on a SAN storage for online retrieval – 4 TB
- Ingested on to NICA – Digital Asset Management System
- For web access an exclusive NICA system is being planned

# Digitisation -- Status

- Business Line – 28, Jan 1994 till date
- The Hindu – 1878 till date (with some intervening missing files)
- Frontline – from Dec 1984 till date
- Sport & Pastime and Sportstar – all issues (in diff layouts)
- Annual Publications – current period in digital form, earlier ones being digitised
- Photographs – current photos – all in digital form – stored in NICA & selected items are hosted on the Net. ([www.thehinduimages.com](http://www.thehinduimages.com))
- Old photos -- 1,00,000+ scanned and about 5,00,000+ yet to be scanned
- Efforts are on to offer the Archives on the Web – similar to our [ePaper](#)

# Digitisation – Demo & Interesting items

- Sport & Pastime
- Sportstar – 1<sup>st</sup> issue (tabloid on newsprint)
- Sportstar -- in A4 Book format
- Sportstar – redesigned & current (tabloid)
- Frontline – 1<sup>st</sup> issue
- Frontline – current Issue
- The Hindu – 1<sup>st</sup> Jan 1963
- The Hindu – 18<sup>th</sup> Jan 2008
- Business Line – 28<sup>th</sup> Jan 1994 – 1<sup>st</sup> Issue
- Business Line – 18<sup>th</sup> Jan 2008
- [Interesting Items](#)