



P@NOPTIC

A panoptic view of Databases, XML and IR

David Hawking
CSIRO ICT Centre, Canberra, Australia
david.hawking@acm.org

Panoptic - Seeing the whole at one view.

Sheffield, 29 Jul 2004

Starting Points

Background

- Goal (1998): Turn PADRE into an Enterprise Search product
- Develop an efficient, practical model of Text + Metadata/Structure
- es.csiro.au/pubs/hawking_tr00b.pdf
- Implement other components, "productize"

Panoptic

- 23 July 1999 - First site
- 23 July 2004 - 34 commercial sites, incl. 1 Canadian, 1 UK
- Databases – email – continuous media – events

Lessons learned

- The customer is always right BUT no two customers are the same!
- The publisher, NOT the searcher, is the customer.
- Content doesn't cut it
- TRUE and FALSE are enough

An example from an oil exploration company

Drilling information recorded in many forms

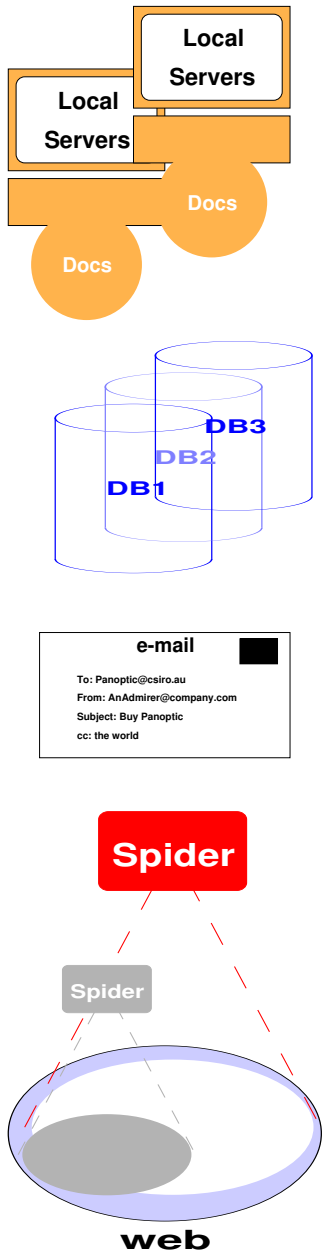
- email
- word processing documents
- spreadsheets
- database records
- OCR-ed faxes

Inconsistent vocabulary

- e.g. "stuck pipe" represented in many ways

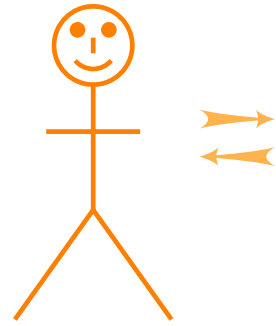
Results are wells, not documents

Result display is tabular - links to GENESIS application



1. Gathering

2. Extracting



3. Indexing



4. Querying

6. Presenting

5. Ranking

Key Issues

1. Handling heterogeneity
2. Answer value v. answer relevance
3. Index the database or webpages derived from it?
4. Answer extent
5. Searching for services
6. Boundary between IR & DB

1. Handling Heterogeneity

“Metadata” classes

Example query element: "author=Baeza-Yates"

- (23,700 Google results for Baeza-Yates)
- author information represented in many different ways:
 - * `<meta name=author content="Arjen de Vries">`
 - * `dc.creator`, `dc.contributor`, `dc.author`, ...
 - * From: `mounia@cs.qmul.ac.uk` (Mounia Lalmas)
 - * `/article/fm/au`
 - * AUTH field in database
 - * perhaps external to document

Panoptic maps author, title, publisher, language etc to **metadata classes**, represented by single letters. i.e. `a:Baeza-Yates`

Raises the lowest common denominator for search over combination of web-pages, email archives, XML collections, databases. **Find Baeza-Yates in `/article/fm/au`** is no good in this context.

Heterogeneous Gathering

Crawl web pages

- extraction programs for storing metadata from non-HTML
- crawl XML
- external metadata tagging based on URL prefixes

Use db2xml plus JDBC bridge to convert DB records to XML files

- SQL query defines the records/fields to be indexed [sals] `select * from employees e, salaries s where e.employeeid = s.employeeid`
- sals becomes element name – can map to metadata classes
- See www.panopticsearch.com/AdminHelp/database.html

Filecopying – with text and metadata extraction

Email – scanners for different mail formats

Administration of Database Indexing

Location Edit View Go Bookmarks Tools Settings Window Help

Location: <https://kangaroo.panopticssearch.com/search/admin/database-collection.cgi>

Administration (Add Database Collection) [Home](#) | [Help](#)

Add Database Collection

This page allows you to add a database collection.

Collection [Help](#)

Display Name: Internal Name:
Admin Email:

Database [Help](#)

Database url: Driver:
Username: Password:

Information Selection [Help](#)

SQL Query: [All Tables](#)

Security [Help](#)

Access Restriction:

Panoptic query elements

Words, phrases, wildcards

- Workshop on databases and IR
- "to be or not to be"
- David.Hawking@csiro.au
- "data* and *formation ret*"

Metadata elements

- a:maarek p:ibm d>feb
- a:*aa*

Query operators

Boolean is too difficult for many

- cat AND dog - find me documents with cat and documents with dog!
- default to AND (with fallback), use brackets for alternatives
- [a:Rölleke a:Roelleke] - disjunction group

negation: !

mandatory exclusion: -

mandatory inclusion: +

- +u:ibm.il [a:Maarek a:Sofer] xml databases !"technical report"

Constraint Counting and Result Tiers

Each query element (dysjunction group is a single element) represents a constraint.

- `+u:ibm.il [a:maarek a:sofer] xml databases !"technical report"`
 - * five constraints - one mandatory

Primary results sort is number of constraints matched.

Results are presented in tiers:

- fully matching documents
- documents matching n-1 constraints out of n

Fully matching tier represents AND – all tiers constitute OR

Dropped both MAYBE state and constraint ordering:

- Sub-tiers caused confusion
- MAYBE almost always equivalent to FALSE

Added scoping operators:– mandatory in/exclusion with no constraint.

As deployed at NRC

Location Edit View Go Bookmarks Tools Settings Window Help

Location: http://index.nrc-cnrc.gc.ca/cgi-bin/rp-ps.cgi?collection=rp_pubs&form=rp_e

National Research Council Canada Conseil national de recherches Canada

Français	Contact Us	Help	Search	Canada Site
NRC Research Press Home	What's New	CISTI Home	NRC Home	

NRC
Research Press

Celebrating
75 years

Journals ▶
Affiliated journals ▶
Monographs ▶
Conference Proceedings ▶
Customer service ▶
Publishing services ▶

Search Results

Anywhere:

Author:
To search for a full name, use quotation marks and begin with the last name ("Smith John A"). Use a wildcard if unsure of full name or initials ("Smith John*" or "Smith J*").

Title:

Abstract:

Published: After (exclusive) Before (exclusive)

Volume: **Pages:** (enter first or last page number)

Journal:
Use CTRL-click for multiple selections

Condensed results (no summaries):

NRC-CNRC

Date modified: 2004-07-17 [Important Notices](#)

Most people use very simple queries or a form.

Result presentation in tiers

Location Edit View Go Bookmarks Tools Settings Window Help

Location: <http://www.panopticsearch.com/search/search.cgi?query=a%3AHawking+network+computing+%21ninemsn&collection=panoptic-commercial&form=advanced>

Searching P@NOPTIC Commercial Site [Updated: Jul 18 2004] Form: advanced simple

P@NOPTIC SEARCH a:Hawking network computing lninemsn search

[Query: a:Hawking network computing lninemsn -- Refine this query Documents: 1 fully matching plus 79 partially matching]

Fully matching documents

1. Panoptic: Enterprise Search Engine

Summary: ... Panoptic Wins **Network Computing's** Editor's Choice Award Panoptic Wins **Network Computing's** Well ... Editor's Choice Award Panoptic Wins **Network Computing's** Well Connected Award Panoptic v 5. 1. ...

Author: David Hawking

Description: Home page for Panoptic Enterprise Search Engine

Keywords: Intranet, Enterprise, Search, Panoptic, CSIRO, panoptic

Publisher: CSIRO Mathematical and Information Sciences

<http://www.panopticsearch.com/> - 15k - [html] - **Cached:** 1 - 5 nov 2003

Documents matching 3 out of 4 search constraints

2. Panoptic: Panoptic Customers

Summary: ... Australian Partnership for Advanced **Computing** Civil Aviation Safety Authority COMSUPER Operated by ... **NINEMSN** Channel 9 and Microsoft **Network** Research Finder Search All Australian Research ...

Author: David Hawking

Description: List of Customers of the Panoptic Enterprise Search Engine

Keywords: Intranet, Enterprise, Search, Panoptic, CSIRO, panoptic

Publisher: CSIRO Mathematical and Information Sciences

<http://www.panopticsearch.com/Products/try-us.html> - 17k - [html] - **Cached:** 2 - 30 jun 2004

3. Panoptic: Get a Panoptic Quote

Summary: ... 50000-100000 100001-250000 250001-500000 500001-1000000 1000001-2500000 2500001-5000000 5000001-10000000 Location of Documents: Please Select Internal **Network** Intranet External Web Site Both Internal and External Organisation Details Organisation ...

Author: David Hawking

Description: Get a quote for Panoptic

Keywords: Enterprise, Search, Intranet, Panoptic, Price, Quote

Publisher: CSIRO Mathematical and Information Sciences

<http://www.panopticsearch.com/Products/quote.html> - 12k - [html] - **Cached:** 3 - 21 nov 2003

4. Panoptic: Backing Up

Summary: ... However, if your organisation has many gigabytes of data, it takes a long time to re-crawl and re-index the pages. ... In principle, the data on a Panoptic search engine doesn't need backing up because it is only a copy of data stored elsewhere on your ...

Author: David Hawking

Description: How to backup a Panoptic Enterprise Search Engine

Keywords: Enterprise, Search, Panoptic, CSIRO, administration, backup

emacs Kons Kons 1 2 3

A distributed approach

Search a database and the web pages, then merge results

The screenshot shows a web browser window displaying search results for "drivers licence" on the Queensland Government website. The browser's address bar shows the URL "http://www.search.qld.gov.au/TextSearch". The website header includes the Queensland Government logo and navigation links like "Home", "Site map", and "Contact us". Below the header, there are navigation tabs for "Services for Queenslanders", "Business & Industry", "Leisure & culture", "Find a service", and "Departments".

The search results are displayed in a list format. Each result includes an icon (star for top results, information for others), a title, a percentage, and a brief description. The top result is "Open Drivers Licence" (100%), followed by "Page : Inventory report - Queensland Transport" (100%), "New Queensland driver licence home" (96%), "QT - Keeping your licence home" (94%), "Provisional Drivers Licence" (93%), "Learner drivers" (92%), "QT - Licence demerit points home" (91%), "Snapshot of the licence project" (89%), and another "Snapshot of the licence project" (88%).

On the right side of the search results, there are two sections: "Narrow Search" and "Widen Search". The "Narrow Search" section lists related terms such as "renewal of drivers licence", "vehicle drivers licence", "renewal of accredited driver licence", "open drivers licence", "provisional drivers licence", "pilot drivers licence", "queensland driver licence", "replacement drivers licence", "escort drivers licence", "new driver licence", "driver trainer licence", and "driver licence home". The "Widen Search" section lists broader terms like "driver crashes", "driver licence", "young drivers", "vehicle drivers", "renewal of drivers", "provisional licence", "learner drivers", "driver trainer", "driver education", "renewal of accredited drivers", "heavy drivers", "open drivers", "open licence", "provisional drivers", "drivers certificate", "pilot drivers", "learner licence", "plates drivers", "drivers log", and "truck drivers".

The browser's taskbar at the bottom shows several open applications, including a file explorer and a terminal window.

Thoughts (on heterogeneity)

Started with an IR engine

- Could start with a database engine

Usually deploy in "centralised" mode

- all data in a single index or meta-index
- distributed is a possible alternative

Are some types of result better than others?

- e.g Web pages better than database records?

Structural mapping OK - what about vocabulary mapping?

- Thesaurus
- Metadata-aware thesaurus?
 - * baker, smith, ...

2. Value v. Relevance

A lesson from Web Search

Web searchers want valuable resources, not individual relevant documents

– NASA homepage rather than thousands of individual NASA documents

In webs you can tell what the valuable resources are from anchor text, links and URL structure.

Example 1

Location Edit View Go Bookmarks Tools Settings Window Help

Location: <http://www.google.com.au/search?q=information+retrieval&ie=ISO-8859-1&hl=en&btnG=Google+Search&meta=>

Google Search [Advanced Search](#) [Preferences](#)

Search: the web pages from Australia

Web Results 1 - 100 of about 3,310,000 for **information retrieval** [definition]. (0.53 seconds)

INFORMATION RETRIEVAL
INFORMATION RETRIEVAL. A book by. ... **Information Retrieval** Group, University of Glasgow. PREFACE TO THE SECOND EDITION (London: Butterworths, 1979). ...
www.dcs.gla.ac.uk/Keith/Preface.html - 7k - [Cached](#) - [Similar pages](#)

Information Retrieval
INFORMATION RETRIEVAL. Second Edition ... **Information Retrieval** Group Department of Computing Science University of Glasgow Glasgow G12 8QQ SCOTLAND. ...
www.dcs.gla.ac.uk/~iain/keith/ - 5k - [Cached](#) - [Similar pages](#)
[[More results from www.dcs.gla.ac.uk](#)]

Modern Information Retrieval
(back cover). Modern **Information Retrieval**. ... Inc. You can order this book on-line with a secure form, or search other titles from A/W about **Information Retrieval**. ...
www.sims.berkeley.edu/~hears/rbook/ - 9k - [Cached](#) - [Similar pages](#)

Modern Information Retrieval - Glossary
Modern **Information Retrieval** ... Distributed **information retrieval**. The application of distributed computing techniques to solve **information retrieval** problems. ...
www.sims.berkeley.edu/~hears/rbook/glossary.html - 40k - [Cached](#) - [Similar pages](#)
[[More results from www.sims.berkeley.edu](#)]

UMASS Amherst: Center for Intelligent Information Retrieval
The Center for Intelligent **Information Retrieval**, a National Science Foundation-created S/UCRC Center, is one of the leading **information retrieval** research ...
ciir.cs.umass.edu/ - 6k - [Cached](#) - [Similar pages](#)

International Symposium on Music Information Retrieval (MUSIC IR ...
ciir.cs.umass.edu/music2000/ - 1k - [Cached](#) - [Similar pages](#)
[[More results from ciir.cs.umass.edu](#)]

Information Retrieval Research - SearchTools Topics
Research on **Information Retrieval**, Library Experience and other academic work related to web site search tools. ... **Information Retrieval** Research. ...
www.searchools.com/info/info-retrieval.html - 22k - [Cached](#) - [Similar pages](#)

The Information Retrieval in Chemistry
The **Information Retrieval** in Chemistry. WWW Server. ... Last full-scale update: Jan. 5, 2004 Copyright: The **Information Retrieval** in Chemistry Team. ...

Sponsored Links
Automated Report Delivery
Schedule and distribute **information** via the Web or e-mail with WebFOCUS
<http://www.informationbuilders.com/>

Information Retrieval
semantic-based IR/IE on natural language texts; 17 languages
www.petame.com

Discover The Essence
Get rid of redundant reading!
Intelligent text mining software
www.texRay.com
[See your message here...](#)

Example 2a

Searching Test External (No Duplicates) [Updated: May 30 2003] Form: [simple](#) [advanced](#)

**P@NOPTIC
SEARCH**

[Query: library -- Documents: 22751 fully matching plus 0 partially matching]

Fully matching documents

- 1 100 [The Library of the Australian National University](#)**
... Science. Astronomy **Library**. Chemistry **Library**. Earth Sciences **Library**. ... Eccles Medical Sciences **Library**. Physics **Library**. N. Aus. ...
<http://anulib.anu.edu.au/> - 24k - [Cached](#) - 7 dec 2001
- 2 99 [ANU Library Catalogue](#)**
... Welcome to the ANU **Library** Catalogue. Search the **Library** Catalogue. ... 50. Search other **library** catalogues Go to the ANU **Library** ...
<http://library.anu.edu.au/> - 10k - [Cached](#) - 19 nov 2001
- 3 87 [ANU Library Intranet - ANU Library Mailing Lists Archives](#)**
...
<http://info.anu.edu.au/> - 13k - [Cached](#) - No Date
- 4 81 [ANU Library: Science - Science and Technology Collection](#)**
... 6125 2551 Astronomy Branch **Library**. Location. The Astronomy **Library** is housed at two sites. ... Locating Items in the **Library**. You can search the ANU **Library** Catalogue, INNOPAC in ...
<http://anulib.anu.edu.au/clusters/science/collection/> - 19k - [Cached](#) - 15 jun 2002
- 5 77 [ANU Library - About the Library](#)**
... Publications and Documentation. SIS-**Library** Committees. Opening Hours. ... Site Map. SIS-**Library** Contact Points. Mission Statement. ... 1999 ANU **Library** Maintained by. Elect. Publishing Marketing Unit. ...
<http://anulib.anu.edu.au/about/> - 8k - [Cached](#) - 18 oct 2002
- 6 76 [Asian Studies WWW VL](#)**
... Asian Studies WWW Virtual **Library**. Established. 24 March 1994. ... Asia-Pacific **Library** Conferences. conferences of interest to librarians. ...
<http://coombs.anu.edu.au/WWWVL-AsianStudies.html> - 24k - [Cached](#) - 5 dec 2001
- 7 76 [ANU Institute of the Arts Library & Resource Centre](#)**
Contact. Institute of the Arts Librarian. Comments and suggestions. Original URL. <http://www.anu.edu.au/ITA-AUSARTS>. Last updated 6 July 2001 ...
<http://www.anu.edu.au/ITA/AusArts/> - 6k - [Cached](#) - 1 aug 1994

Document: Done (0.268 secs)

Example 2b

Location Edit View Go Bookmarks Tools Settings Window Help
Location: http://peace.anu.edu.au/search/search.cgi?query=library&collection=test-external-no-dupes

Searching Test External (No Duplicates) [Updated: May 30 2003] Form: simple advanced

P@NOPTIC SEARCH
library search

[Query: library -- Documents: 22063 fully matching plus 0 partially matching]

Fully matching documents

- 1 100 untitled**
... Astronomy-related Libraries Aarhus University, Institute of Physics Astronomy **Library** Anglo-Australian Observatory **Library** AAO **Library** Astronomical Observatory of Trieste **Library** OAT Astronomische Nachrichten Astronomical Notices News in Astronomy and Astrophysics Astronomy Librarians and Libraries Addresses Baja Astronomical Observatory **Library** Bologna University ... Wolbach **Library** Harvard-CFA **Library** of the Astronomical ...
http://msowww.anu.edu.au/astronomy/astroweb/astro_library.html - 10k - **Cached** - 31 oct 2001
- 2 98 ANU Library - Loans Schedules**
... the **library** only Use within. the **library** only Bound Serials in ITA **Library**. ... the **library** only Use within. the **library** only Reserve Items except ITA **Library**. ... the **library** only Audio Visual material except ITA **Library** 2 days Videos use within the **library** only Audio tapes 2 days Videos use. ...
<http://anulib.anu.edu.au/loans/loans.html> - 23k - **Cached** - 20 sep 2001
- 3 98 The Australian National University Library - Library Committees**
... LAWLAC Law **Library** Advisory Committee. Membership and Minutes SCILAC-Science **Library** Advisory Committee. ... ESLAC-Earth Sciences **Library** Advisory Committee. MATHLAC Mathematical Sciences **Library** Advisory Committee PHYSLAC ...
<http://anulib.anu.edu.au/about/libcommittees.html> - 11k - **Cached** - 26 sep 2002
- 4 98 Tru64 UNIX**
... **library** function. puts. **library** function. Q. qsort. **library** function. ... **library** function. srand. **library** function. scanf. **library** function. ... strchr. **library** function. strcmp. **library** function. strcoll. **library** function. ...
http://mf.apac.edu.au/facilities/sc/compaq_mirror/HTML/ARH9NCTE/DOC0033.HTM - 36k - **Cached** - 28 apr 2001
- 5 98 Biology cancellations 1996**
... Paul MCNAMARA, **Library**. BSBI ABSTRACTS. Hancock. Paul MCNAMARA, **Library**. ... Hancock. Paul MCNAMARA, **Library**. GEDDERMA. Hancock. Paul MCNAMARA, **Library**. ... Hancock. Paul MCNAMARA, **Library**. IAWA JOURNAL. Hancock. Paul MCNAMARA, **Library**. ...
<http://anulib.anu.edu.au/clusters/science/cancellations/biology1996.html> - 21k - **Cached** - 16 jan 2001
- 6 98 Architecture of our Century**
... 1928. BIG. Asplund G. **Library**. Stockholm. Stockholm City **Library**. ... BIG. Bunning Madden. **Library**. Canberra. National **Library** of Australia. ... Bunning Madden. **Library**. Canberra. National **Library** of Australia. ...
<http://anulib.anu.edu.au/Access/Arch/Arch/Arch/Arch/200109.html> - 4k - **Cached** - 1 oct 1999

Library of Australia

Example 3

The screenshot shows a web browser window with the following content:

Browser Title: Panoptic Comparison - Konqueror

Address Bar: http://tt/search/compare-searches.cgi?query=events&collection=asx

Page Header: Searching Australian Stock Exchange (ASX) [Updated: May 7 2004] Form: simple advanced

Search Results:

- 1 100 ASX - Upcoming Events**
... Upcoming **events** On this page you will find information on: Channel Nine Money Expo 28-30 May Traders Investors Expo Melbourne 14 15 May ASX World Link and Singapore Exchange Seminar Better ...
http://www.asx.com.au/UpcomingEvents_HP2.shtm - 26k - **Cached** - No Date
- 2 99 Calendar of Events**
... Calendar of **Events** Event/Conference Date Location More information Resnik WA Financial Planning Conference 24th 25th February 2003 Perth www. resnik. com. au Resnik Tax Strategies Conference 5th 7th March 2003 ...
http://www.asx.com.au/about/13/AdviserServicesCalendarofevents_AA3.shtm - 13k - **Cached** - No Date
- 3 67 Register for MyASX**
... Tick box to receive ASX Trader's Update view example Strategies and news from ASX's options, warrants, and futures markets. ... You may unsubscribe at any time. Tick box to receive ASX Investor Update view example Occasional emails on ASX products, courses, discounts and upcoming **events**. ... Please tick the box if you would like to receive updates. To receive information about **events** in your area, please ensure your postcode is correct. ...
<http://www.asx.com.au/asx/myasx/register.do> - 33k - **Cached** - No Date
- 4 66 ASX - Australian Stock Exchange**
... immediately. ASX Investor Education refunds, transfers and cancellations policy ASX routinely conducts investor education **events** and online courses, and in certain circumstances tickets for these **events** or courses may be purchased by ...
<http://www.asx.com.au/legal/GeneralConditions.shtm> - 27k - **Cached** - No Date

Right Side Navigation:

- Getting started (Find out about...)
- Free annual reports
- HOME, MARKET STATISTICS, COMPANY RESEARCH, ASX MARKETS, ASX SHAREHOLDER INFORMATION, FLOAT
- Search www.asx.com.au (Enter your search word or phrase below: events)
- Search results (Documents 1 to 10 of 300, matching the query "events".)
- chapter 7 - timetables (Appendix 3A Timetable The days indicated in the "Business day" the event in the minimum (or maximum) time limit specified for the ASX agrees to quote an entity's securities on a "cum" then "ex" it announces a +record date, the <http://www.asx.com.au/listingrules/appendices/app3a.doc>)
- Prospectus summary - TOLL Holdings Limited (TOLL Holdings Limited The information below is a summary only acquaint themselves with the full details in the prospectus or contact your advisor.ASX code:. TOLG. Quoted date:. 18 April 2001. Issue type subordinated Notes. Share registry:. http://www.asx.com.au/markets/14/irmdescriptiontolg_am4.shtm)
- ASX - Australian Stock Exchange site map (The www.asx.com.au site map details each of the sections of the website. <http://www.asx.com.au/sitemap.shtm>)
- Prospectus summary - Leighton Holdings Limited (Leighton Holdings Limited The information below is a summary or should acquaint themselves with the full details in the prospectus advisor.ASX code:. LEIGA. Quotation date:. 10 December 2003.)

A challenge for DB/XML retrieval

How do you get the Web search win in collections of XML documents or in a database, or in email?

- Exploiting references - do they mean the same as on the web?
- Is there an equivalent of search and browse?
- How to suppress repetition?
 - * Google suppresses results more than two results from the same site
 - * Panoptic implements an increasing boredom factor.

3. Index web pages OR the DB records from which they were generated?

Searching the DB

can you map DB records to useful URLs?

just display the content of the record?

xsltproc (e.g. with Genesis links)

how do you include database results among web pages?

how do you rank DB records relative to web pages?

some customers store HTML or XML in their DB records.

– SE mustn't trash it, but it can trash results page.

Searching generated Web pages

possible multiple duplication/repetition

do links automatically generated from DB records mean anything?

will external people link to automatically generated pages?

– depends upon URLs

4. Answer Extent

Parallel between web search and INEX 'exhaustivity' / 'specificity'

XML blurs the definition of "document"

- So does hypertext

Keith van R's IR book is a good example.

A relevant website is a better result than a web page

- right level, can avoid repetition
- browse
- local search

Visual blocks within web pages.

5. Searching for services

Currently many people search the web for services

- buy a ticket to Sheffield
- download MP3 of Waltzing Matilda
- renew my staff card

What will be the effect of SOAP/ Web Services

- Does it work in a heterogeneous environment
- Exploiting links, anchortext etc.

6. Boundary between IR & DB

How far should an IR system go?

- Joins, set computations
- Are databases really a special case of IR systems?

7. Some Applications

Event Search prototype

Location Edit View Go Bookmarks Tools Settings Window Help
Location: http://localhost.localdomain/search/padre-sw.cgi?query=%5Bben+%22communities+online%22+zoo%5D++%

Search: Events

P@NOPTIC SEARCH [ben "communities online" zoo] % w>20001231 search

[Query: [ben "communities online" zoo] % w>20001231 -- Events.]

18 oct 2002

1. [RMIT keynote preso at Zoo](#)
Need I say more. ...
2. [PADRE mods for Communities Online](#)
PADRE mods for **Communities Online**. Real deadline is Wednesday 22nd but I'll be in Melbourne on 21st. ...

19 oct 2002

3. [RMIT keynote preso at Zoo](#)
Need I say more. ...
4. [PADRE mods for Communities Online](#)
PADRE mods for **Communities Online**. Real deadline is Wednesday 22nd but I'll be in Melbourne on 21st. ...

20 oct 2002

5. [Pan5 CD for Ben Tan](#)
As promised by Dave ...
6. [RMIT keynote preso at Zoo](#)
Need I say more. ...
7. [PADRE mods for Communities Online](#)

CMWeb Search prototype

Slide 2

Searching ABC CMWeb [Updated: Sep 26 2003]

P@NOPTIC
SEARCH

terrorism

search

All the words:

The phrase:

Any of the words:

Title:

Author:

Location:

News Service:

Subject:

Published after:

Format:

Published before:

[Query: terrorism -- Documents: 15 fully matching plus 0 partially matching]

Fully matching documents

1. [Labor pursues Howard over Iraq terrorism claims](#) ABC TV News [00:00:00 - 00:00:38]



Location: Australia

See also:

[Video](#) | [original media file](#) | [whole media file](#) | [more information](#) - 18 sep 2003

2. [Labor pursues Howard over Iraq terrorism claims](#) Lateline [00:00:00 - 00:01:41]

People Finder prototype

Slide 4

CSIRO Expert Search are brought to you by the [the TED group](#)

Contacts details for **David Hawking**

No Image Present	David Hawking
	Contact Details CSIRO Mathematical and Information Sciences Street Address: Building 108, North Road, ANU Campus, Acton, ACT 2601, Australia Postal Address: GPO Box 664, Canberra, ACT 2601, Australia E-mail: David.Hawking@cmis.csiro.au Telephone: (02) 6216 7060 Fax: (02) 6216 7111

Other Names associated with **panoptic**

Experts associated with: panoptic				
Name	Phone	Fax	Email	Evidence
David Hawking	(02) 6216 7060	(02) 6216 7111	David.Hawking@cmis.csiro.au	more details and evidence
Nick Craswell	(02)6216 7059	(02)6216 7111	Nick.Craswell@cmis.csiro.au	more details and evidence
Graham Williams	(02) 6216 7042	(02) 6216 7111	Graham.Williams@cmis.csiro.au	more details and evidence

Documents Related to **David Hawking**

Searching CMIS [Updated: Oct 17 2003] Form: [simple](#) [advanced](#)

Panoptic Search

[Query: +[panoptic] +[\"David Hawking\"] -- Documents: 15 fully matching plus 0 partially matching]

Recap of Key Issues

1. Handling heterogeneity
2. Answer value v. answer relevance
3. Index the database or webpages derived from it?
4. Answer extent
5. Searching for services
6. Boundary between IR & DB

Where to from here?

What the world needs now ...

Is a heterogeneous enterprise search collection with real queries

- intranet pages
- website pages
- databases
- email
- fileshares
- XML

Possibilities

- W3C collection
 - * public web and email (real queries)
- A bankrupt company
 - * Enron? but what queries?

David.Hawking@acm.org

es.csiro.au – thylacine.panopticsearch.com/dave/irdbxml.pdf