



CSA Illustrata, Gale Virtual Reference Library, and Cambridge Journals



CSA Illustrata
is a giant leap in
scholarly information
retrieval from
commercially
available systems.

One of my picks is the innovative, cutting-edge CSA Illustrata database, launched at Midwinter ALA, with nearly 1 million illustrations extracted from more than 165,000 scholarly articles. Enhanced by a variety of illustration-specific index terms that describe important attributes of the graphs, charts, photographs, tables, and other figures, it offers unprecedented precision in finding articles that discuss *and* show, say, the phenomenon and management of beach erosion. CSA Illustrata is not merely an evolutionary step or an experimental system of a grant project, but it is a giant leap in scholarly information retrieval from commercially available systems.

The other pick is the Gale Virtual Reference Library (GVRL), which offers several hundreds of encyclopedias and other reference books, published by Gale and at least 30 other publishers such as SAGE, Wiley, and Springer. It includes many award-winning, highly respected, and brand new encyclopedias, such as the *Encyclopaedia Judaica*, which recently won the Dartmouth Medal. Librarians can create their own mix of encyclopedias and keep them forever through Gale's perpetual licensing option. This approach combines the advantages of digital access to every word in the reference works, the complementary functions of HTML and the traditional page layout through PDF, and the peace of mind of functional ownership of acquired resources.

The pan is the Cambridge Journals Online digital collection. The respected content of its 200-plus journals is disgraced by careless software implementation. It is especially troubling that the problems I spotted more than a year ago have not been fixed, and there is no acknowledgement of, let alone corrective actions for, error reports.



the picks

CSA ILLUSTRATA

CSA Illustrata (<http://info.csa.com/csainustrata>) is a giant leap in an era when even the most affluent college and special libraries have to put up with many expensive information-retrieval systems that completely omit, or remain silent about, informative graphic materials in their full-text records. Sometimes they merely indicate the omission with a short note, or they provide only the text of the

1. **Validity of North Shore, Oahu, Hawaiian Islands Surf Observations**
 Caldwell, PC
 Journal of Coastal Research [J. Coast. Res.], Vol. 21, no. 6, pp. 1127-1138, Nov 2005.
 Surf information is imperative for safety, coastal planning, and engineering applications. Daily surf observations made primarily by lifeguards along the north shore of Oahu, Hawaii, have been digitized for the 36-year period from 1968 to 2002. The ...

2. **Artificial Neural Network Classification of Sand in all Visible Submarine and Subaerial Regions of a Digital Image**
 Conner, CL; Fletcher, CH; Barbee, M
 Journal of Coastal Research [J. Coast. Res.], Vol. 21, no. 6, pp. 1173-1177, Nov 2005.
 Factors controlling the distribution of shelf sand as a resource, a component of reef ecosystems, and a dynamic substrate are poorly understood. An initial step in understanding sand accumulation in each of these roles is to identify its areal extent ...

3. **A Risk Assessment for Pacific Leatherback turtles (*Dermochelys coriacea*)**
 Kaplan, JC
 Canadian Journal of Fisheries and Aquatic Sciences [Can. J. Fish. Aquat. Sci.], Vol. 62, no. 8, pp. 1710-1719, Aug 2005.
 Leatherback turtles (*Dermochelys coriacea*) are critically endangered in the eastern and western Pacific Ocean. Here, I estimate the magnitude of two likely causes of their decline: (i) bycatch by longline fishing vessels and (ii) coastal sources of ...

Databases: CSA Illustrata: Natural Sciences
 Descriptors: Coastal zone management | Coastal accretion | Wave processes on beaches | Surf | Wave heights | Recreational planning | Maps

Databases: CSA Illustrata: Natural Sciences
 Descriptors: Sedimentation | Sand | Sedimentation | Sediment transport | Coral | Limnology | Tachyarrhythmia | Maps

Databases: CSA Illustrata: Natural Sciences
 Descriptors: Habitat | Risk assessment | Oceans | Fisheries | Crustaceans | Eggs | Maps

Results list with thumbnails of the illustrations

captions (without making a field-specific index of them). At best, which isn't all that good, they reproduce, in rather clumsy ASCII (American Standard Code for Information Interchange) format, some of the tables, some of the time.

I tested the beta version of CSA Illustrata: Natural Sciences from late December 2006. It not only includes nearly 1 million charts, graphs, photographs, X-ray reproductions, tables, and other illustrative materials from more than 165,000 scholarly articles published in about 100 science journals (many from Blackwell Publishing, CSA's development partner), but it also makes them eminently searchable through exhaustive indexing—or deep indexing, to use CSA's less ambiguous term. CSA plans to add 1 million images a year for

the next 2 years so that by 2009, there will be 3 million searchable images.

Beyond indexing the text of the captions of illustrations, all illustrations (objects) are enhanced and indexed by descriptors, subject terms, and digital object identifiers. Where applicable, CSA Illustrata adds taxonomic, geographic, and statistical terms to refine the indexing of the objects. Each is searchable, in addition to the traditional indexes of author, title, abstract, subject descriptor, and journal name. The object categories and the object descriptors are also browsable.

The results lists show tiny thumbnails of the illustrations. The ones that match the query term(s) have a red frame. Mousing over the thumbnails will show part of the caption.

Larger images are shown when records are selected for preview. Mousing over the images will show a metadata "cloud" with the full caption and the object-specific descriptors; geographic, taxonomic, and statistical terms (as applicable); along with the category names of the illustrations. There are thousands of object descriptors (obviously, the indexing experience and resources of CSA are deployed for this) and more than 50 category names to specify the type of illustrations, such as histograms, scatter plots, study site map, chemical structure, and time series. I wish the statistical and taxonomic terms were also browsable. Having found the pertinent articles, the full-size illustrations and their abstracting and indexing records can be displayed and saved. There are links to the digital collections of the publishers of the journals as well as to aggregators, such as EBSCO's Academic Search Premier, to locate the complete documents.

CSA's pricing for Illustrata starts this year at \$14,900, going up to \$15,600 in 2008 and \$16,500 in 2009. New subscribers in 2008, however, have subscriptions starting at \$18,100, going up to \$19,000 in 2009. Those who wait to subscribe until 2009 will see a starting price of \$21,900. Adjustments will be made to reflect current subscriptions to CSA products.

My short review here, with only two illustrations, obviously can't do justice to the richness of CSA Illustrata. An in-depth open access review with dozens of screen shots appears at my Digital Reference Shelf page (www.gale.com/reference/peter/index.htm).

GALE VIRTUAL REFERENCE LIBRARY

Thomson Gale (which hosts my Digital Reference Shelf review column) introduced GVRL in 2004 with about 300 sources (www.gale.com/ebooks). In the past 3 years, the number of sources in GVRL tripled, depending on how you count the sources. If a title has two editions, or is part of a huge series such as Contemporary Authors, is it one or two titles? Regardless of your counting methodology, GVRL is by far the largest digital ready reference collection. The only fair comparison would be by the number of words and illustrations—Gale estimates that it will have 1 million pages of online content by the time you read this.

GVRL's particular strength is in the many multivolume encyclopedias, providing much more comprehensive information

Title: **Validity of North Shore, Oahu, Hawaiian Islands Surf Observations**
 Author: **Caldwell, PC**
 Affiliation: Hawaii Liaison Office, National Coastal Data Development Center, National Oceanic and Air Administration, 1000 Pope Road, MSB 316, University of Hawaii at Manoa, Honolulu, HI 96822 patrick.caldwell@noaa.gov
 Source: Journal of Coastal Research [J. Coast. Res.], Vol. 21, no. 6, pp. 1127-1138, Nov 2005.
 Objects:

Figure 1. Figure 2. Figure 3. Figure 4. Figure 5. Figure 6. Figure 7. Figure 8.

Figure 7. Caption: Root-mean-square (RMS) error estimates for various subsets versus the observation height (Hsf = Hawaii scale feet). A linear best fit for all subsets is overlaid.
 Category: **Figure**; **Graph**; **Line Graph**
 Object Subject Terms: **Observation height**

Medium-sized illustrations with optionally displayed object-specific metadata 'clouds'

- 7. Ancestors: Baltic Cult of Ancestors.** RÕTA MUKTUPÄVELA. *Encyclopedia of Religion*. Ed. Lindsay Jones. Vol. 1. 2nd ed. Detroit: Macmillan Reference USA, 2005. p327-332. 15 vols.
[Full-text](#) | [6 PDF Pages](#) | [About this Publication](#) | [How to Cite](#)

- 8. Ancient Egypt.** *World Religions*. London: Times Books, 2004. p26-31.
[Full-text with Graphics](#) | [6 PDF Pages](#) | [About this Publication](#) | [How to Cite](#)

- 9. Ancient Egypt and the Afterlife.** *Gale Encyclopedia of the Unusual and Unexplained*. Eds. Brad Steiger and Sherry Hanson Steiger. Vol. 1. Detroit: Gale, 2003. p15-22. 3 vols.
[Full-text with Graphics](#) | [8 PDF Pages](#) | [About this Publication](#) | [How to Cite](#)

- 10. Angola.** Inge Brinkman. *Worldmark Encyclopedia of Religious Practices*. Ed. Thomas Riggs. Vol. 2: Countries: A-L. Detroit: Gale, 2006. p22-29. 3 vols.
[Full-text with Graphics](#) | [8 PDF Pages](#) | [About this Publication](#) | [How to Cite](#)

Substantial articles from a variety of encyclopedias on the topic of burial rituals from GVRL

than the short-entry works that dominate NetLibrary Reference Center or Oxford Reference Online. Gale was clever in recognizing that its own reference works do not cover all disciplines. It, therefore, licensed encyclopedias from Macmillan, Philip's, Idea Group, SAGE, Springer, McGraw-Hill, and other publishers for the GVRL collection.

Importantly, each library can choose its GVRL version—and keep it for good. A high school library can buy the relatively slim, single-volume *World Religions*, published by TIME Books in 2004. A small college library might choose the six-volume *World Religions Reference Library*, published in November 2006. The library of a large seminary could choose, along with the above-mentioned generic reference works of religion, all or some of the other encyclopedias focusing on specific religions, such as the second edition of the *New Catholic Encyclopedia*, the *Encyclopedia of Buddhism*, the *Encyclopedia of Islam and the Muslim World*, and the *Encyclopaedia Judaica*. A very large public or university library might opt for hundreds of encyclopedias and other reference works from the GVRL collection, including the ones mentioned above for the most comprehensive coverage.

In my test collection of 940 reference sources, with more than 1 million encyclopedic articles, I could easily check the breadth of almost the entire GVRL collection when running a quite restrictive unidirectional adjacency search (*burial* ritual**). It yielded 192 articles from dozens of sources beyond the religion-specific ones, including the *Encyclopedia of Death and Dying*, *Countries and Their Cultures*, *Africa: An Encyclopedia for Students*, *American History Through Literature*, the *Encyclopedia of Anthropology*, the *Encyclopedia of Food and Culture*, the *Encyclopedia of India*, and the *Encyclopedia of Modern Asia*, to name just a few.

The vast majority of the sources in my test collection were published in the past few years, including 42 in 2007, 243 in 2006, 208 in 2005, 175 in 2004, 93 in 2003, and 69 in 2002. This bests the currency of print reference collections of even the most well-off libraries. Only a few of the reference works would need a new edition for the next few years (the same is

true for print), and the accessibility in every regard (including the 24/7 remote access) is so incredibly superior with the digital editions that they are very much worth the surcharge (usually 10 percent for small libraries and negotiable for others based on full-time employment or population served; discounts apply if librarians purchase print and ebook formats). Articles are available in HTML and PDF, which reproduces the original pages down to the page numbers. Black-and-white and color images enhance the text.

The rich content is matched by a capable software, which can do proximity and positional operations and offers searching in the article title, publication title, publisher name, author name, keyword, and full-text indexes. The keyword index deserves special mention for its smart composition. It has words extracted from the article title, the first 50 words of the article text, and the back-of-the-book index entries, which reflect the judicious decision of human indexers. GVRL is one of those few systems where the captions of the figures can also be searched.

Searches can be limited by target audiences, to documents with images, by document type, and by major disciplinary areas. This allows for retrieval of articles, for example, on only the economic aspects of depression.

Results can be sorted by document and publication title. Publication year would also be a good sort option. Clustering the results set (if more than 20) by back-of-the-book index terms and, in case of a large mix of GVRL sources, by publication title would be useful. There are a variety of options for format and content to download, print, or email articles. GVRL has the best solutions for emailing results. Search results can be "infomarked" and used later to retrieve the items. These infomarks are lengthy and enigmatic, but they can be hidden behind normal texts as URLs, such as the title of the article, in a Webliography or in a reading list. They can also be emailed.

This is not only the largest ready reference collection, one that can be customized to the max, but also the one with the most substantial, long encyclopedic entries and with the option of perpetual access—the functional equivalent of ownership.



the pan

CAMBRIDGE JOURNALS ONLINE

The University of Oxford and the University of Cambridge—often referred to as Oxbridge—have had a rivalry for hundreds of years in many fields, from sciences to sports. In journal publishing, the competition is also fierce, with each publishing more than 200 journals. Cambridge is somewhat ahead in this rivalry (especially if you count the 20 new journals announced to ring in the new year), just as it is in the Boat Race. However, it is not even in the same league as

Your enquiry has returned 4 matches.

relevance ranking search

Search within results New Search

Subscription access
Free access
Trial access
Full text html as well as PDF articles available
Note: Abstract, PDF & HTML open in a new window

Sort by Relevance Results per page 10 Page 1 of 1 Go to page 1 Go to: First Previous Next Last

Save search Save selected items Export citation Add to basket

Article Description	Save / Export / View citation	Add to basket
View selected abstracts on one page	Select all	<input type="checkbox"/>
<input checked="" type="checkbox"/> KIM - a semantic platform for information extraction and retrieval BORISLAV POPOV ATANAS KIRYAKOV DAMYAN OGNYANOFF DIMITAR MANOV ANGEL KIRILOV Natural Language Engineering, Volume 10, Issue 3-4, Sep 2004, pp 375-392 doi: 10.1017/S135132490400347X, Published online by Cambridge University Press 11 Oct 2004 Abstract Add to basket £14.00 / \$20.00	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/> KIM - a semantic platform for information extraction and retrieval BORISLAV POPOV ATANAS KIRYAKOV DAMYAN OGNYANOFF DIMITAR MANOV ANGEL KIRILOV Natural Language Engineering, Volume 10, Issue 3-4, Sep 2004, pp 375-392	<input type="checkbox"/>	<input type="checkbox"/>

Duplicate and triplicate records flood all the results lists in CJO, but unless the set is very small, they are often scattered.

Oxford in offering digital access to journal collections, simply because of its remarkably poor software. The software for Cambridge Journals Online (CJO; <http://journals.cambridge.org>) makes searching the collection of more than 140,000 journal articles as pleasant as a root canal procedure.

For starters, entering a multiple-word query such as **brain damage** triggers an OR operation and returns more than a quarter of the records in the database, with either **brain** or **damage** in the text. It is probably one of the handfuls of systems that did not learn that interpreting a space to be a Boolean OR operator is a bad idea, particularly when coupled with a simplistic relevance-ranking algorithm. The search **brain and damage** yields the same results because the Boolean operator must be in uppercase. Otherwise, it is ignored; hence the OR operator kicks in.

Regardless of the search query, the results list will be full of duplicates, sometimes triplicates. They are not necessarily adjacent in the results list (so much for relevance ranking), but if you want to sort the results by title, you could see them—except when you can't. The problem is that title sort doesn't work reliably—the message that “the system cannot accommodate your request” recurs. In other cases, the title sort reruns the search and produces no results. A search for **Dar fur** produced 401 hits. Sorting the results list by title then yielded zero hits. I have no explanation for this (but I have a screen shot). I'm calling this new type of sort the “cannibal sort.”

Another “interesting feature” of the software is how, in the title list, it adds characters to some words that don't appear in the title of the full record, such as cannibalss and cannibalismism. This dittography is a common problem in data entry by humans, but it is an enigma for me how it can be generated from the correct title fields by the software. This is not the only enigma in CJO.

There are links to CrossRef and Google Scholar to find papers citing the one you are looking at, but be careful with the

reported citedness of Google Scholar. It often inflates its citedness count by phantom citations, as I illustrated in *Online Information Review* (V. 30, No. 3, 2006; <http://projects.ics.hawaii.edu/~jacso/PDFs/jacso-deflated-inflated.pdf>).

There are pretty icons to indicate the type of article—except when they don't. A large number of records on the results lists have no icons whatsoever. Perhaps they are open access articles. No, they aren't; such articles have their own icon, not shown among the other icons.

There are potentially good filters to limit the search to specific journals, to journals held by the library, or by discipline and article type. The latter has some oddities. The type “errata” produces only an error message. The name jurisprudence for article type is rather unusual, and its spelling as Juris Prudence (which would make Justicia—no first name, we are goddesses—pull down her blindfold) guarantees that there will be zero results for any search using this filter. There is no search results clustering by any attributes.

It is possible to export records for bibliographic-management software (using the RIS format), but most of the records contain the HTML formatting codes to indicate emphasized (bold) and small (lowercase) typeface. Some records do not have this anomaly.

```
AU - Lobban, Richard
PY - 2001/Jun/30
TI - E<SMALL>NDRE</SMALL> S<SMALL>TIANSEN AND</SMALL>
M<SMALL>ICHAEL</SMALL> K<SMALL>EVANE</SMALL>,
<SMALL>ED</SMALL>. <EM>Kordofan Invaded: Peripheral I
Social Transformation in Islamic Africa</EM> (Boston:
1998). Pp. 319. $94 cloth.
JF - International Journal of Middle East Studies
KW -
SP - 290
EP - 292
VL - 32
IS - 02
```

Messy export format from Cambridge Journals Online

There is a help file, but it does not help much with statements such as “By placing extra constraints on your search, you can carry out more advanced searches using the same fields and search strings as for simple queries. You do this by using the words AND, OR or NOT.” I have been under the impression for the past 30-plus years that OR does not place an extra constraint on the search; it does just the opposite. In the help file, there is one entry about searching, another about advanced search, and eight about the shopping basket. It might help sales if someone with modest competence would mind the store first and fix at least the most irritating errors, which hinder access to a valuable content.

Péter Jacsó (jacso@hawaii.edu) is professor of library & information science at the University of Hawaii's department of information and computer sciences.

Comments? Email letters to the editor to marydee@xmission.com.