

# Deep Indexing and Discovery of Tables and Figures

Dr. Robert J. Sandusky  
AUL for Information Technology  
Richard J. Daley Library  
University of Illinois at Chicago

NISO Discovery Tools Forum 2008

# Outline

- Deep indexing
- Challenges
- Is there a need?
- Implementations
- Opportunities and dangers
- Issues for research and development

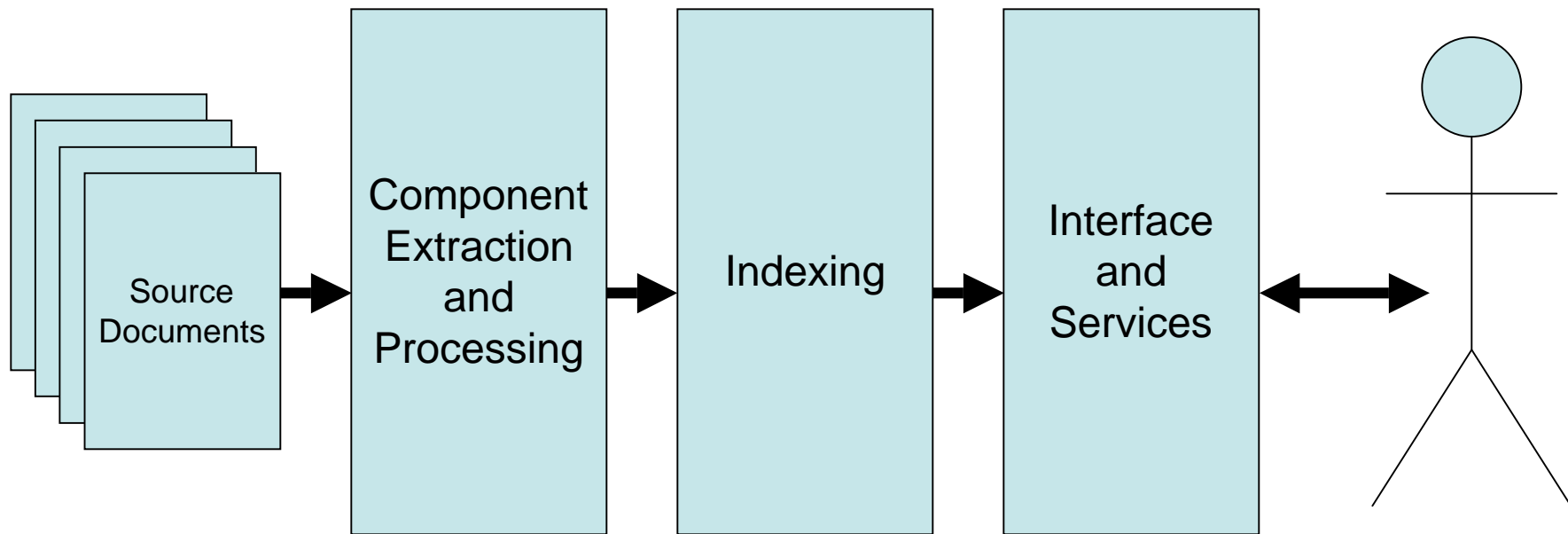
# Deep Indexing

- Working definition:
  - An indexing system that supports discovery of information objects at levels of granularity beyond the abstract or article
- E.g., Figures, tables, other components drawn from scholarly journal articles
  - Make “information contained in tables, figures, and graphs accessible through explicit, detailed indexing” - E. Vawter, *Searcher*, May 2007.

# Deep Indexing

- Is there a need?
  - T&F hold data, images - organisms, objects, phenomena - in a condensed fashion
  - T&F important for judging article level relevance (Bishop, 1999; Sandusky & Tenopir, in press; Olsen, 1994; Stewart, 1996)
  - Users do seek photographs, maps, data with high specificity for teaching and research (Sandusky & Tenopir; Sandusky, Tenopir & Casado)
  - User want more efficient searching
    - When seeking a map or image, be able to find one

# General Model & Challenges



# Challenges

- Extracting & processing components of interest
- Indexing the components
- End-user experience

# Challenges: Extraction & Processing

- Two extraction approaches
  - Identify, extract, and mark up components from untagged documents (e.g., PDF, Word)
  - Obtain tagged components from commercial / association publication processes
- Processing depends upon many factors
  - What do you want to expose?
  - How do you expose it?

# Challenges: Indexing

- Computational indexing approaches
- Human indexing
- Choose elements to index:
  - Captions / legends
  - Content of tables
  - Text referring to tables and figures
  - Images

# Challenges: End-User Experience

- End-user experience
  - Will users recognize they have direct access to components?
    - Will users know how to utilize it? (Bishop)
  - What does it mean to rank components returned in a search set? (Liu, et. al.)
  - What is a good component search interface? (Hearst, et al.)

# Deep Indexing Implementations

- Four implementations
  - DeLiver (Desktop Link to Virtual Engineering Resources) - 1996-2000
  - Illustrata - 2006+
  - TableSeer (CiteSeer) - 2006+
  - BioText Search Engine - 2007+

# Deep Indexing Implementations

- DeLiver (Desktop Link to Virtual Engineering Resources) - 1996-2000
  - Full text of 50 journals in engineering, physics, and computer science
  - SGML feeds directly from publishers, normalized and indexed

# Deep Indexing Implementations


- DeLiver
  - Direct, automatic indexing of components
    - Full text, abstract, title, author, author affiliation, section headings, journal title
    - Article body only
    - Cited references
    - Table captions
    - Table text
    - Figure caption

# Deep Indexing Implementations

- DeLiver
  - “Extended citation”
    - Citation, affiliation, abstract, link to corresponding A&I index records
    - Link to full text
    - List of and links to figures and tables from article
      - Caption and icon (not image thumbnail)
    - References (with links)
    - Links to citing articles in database

Oscillatory Behavior of the Rate of Escape through an Unstable Limit Cycle - Microsoft Internet Explorer

File Edit View Favorites Tools Help

 Desktop Link to Virtual Engineering Resources

---

Maier, Robert S., and Stein, D. L., "Oscillatory Behavior of the Rate of Escape through an Unstable Limit Cycle," *Physical Review Letters* 77 no. 24 (9 December 1996): 4860 - 4863.

FULL TEXT: [XML/HTML](#), [APS PDF](#)  
CITATIONS & OTHER LINKS TO FULL TEXT: [APS Abstract](#) [INSPEC](#)

**AUTHORS:**

**Maier, Robert S.**  
*Department of Mathematics, University of Arizona, Tucson, Arizona 85721*

**Stein, D. L.**  
*Department of Physics, University of Arizona, Tucson, Arizona 85721*

**ABSTRACT:**

Suppose a two-dimensional dynamical system ...

© 1996 The American Physical Society 0031-9007/96/77(24)/4860(4)/\$6.00

**REFERENCES:**

Theory of Noise-Induced Processes in Special Applications, F. Moss and P. V. E. McClintock (Cambridge University Press, 1989).


P. H.änggi, P. Talkner, and M. Borkovec, *Rev. Mod. Phys.* 62, 251 (1990). [INSPEC](#) [AIP Abstract](#)


T. Naeh, M. M. K. I. orek, B. J. Matkowsky, and Z. Schuss, *SIAM J. Appl. Math.* 50, 595 (1990). [INSPEC](#)

R. S. Maier and D. L. Stein, *Phys. Rev. E* 48, 931 (1993). [INSPEC](#) [AIP Abstract](#) [APS Abstract](#)

...

**FIGURES & TABLES:**

 FIG. 1. The unstable limit cycle 6W of the van der Pol model, and the MPEP, which emerges from the attractor 0,0 and spirals into it. The trajectories exiting from W are optimal trajectories that are small perturbations of the MPEP.

 FIG. 2. A Poincaré section. This sketch shows the points n,pn generated by the optimal trajectories passing by some specified point on 6W. The dots are generated by the MPEP, spiraling into 6W. Cf. Figs. 1-3 of Graham and Tel.

**CITING ARTICLES:**

Luchinsky, D. G., Maier, R. S., Mannella, R., et al., "Experiments on Critical Phenomena in a Noisy Exit Problem," *Physical Review Letters* Vol. 79 no. 17 (27 October 1997): 3109 - 3112. [Full Citation & Links to Full Text](#)

Dykman, M. I., Rabitz, H., Smelyanskiy, V. N., et al., "Resonant Directed Diffusion in Nonadiabatically Driven Systems," *Physical Review Letters* Vol. 79 no. 7 (18 August 1997): 1178 - 1181. [Full Citation & Links to Full Text](#)

...

Figures &  
tables



# Deep Indexing Implementations

- Illustrata - 2006+
  - Developed in 2005-2006 by CSA (now ProQuest CSA)
  - Prototype evaluated summer 2006
    - 300,000 figure / table objects
  - 60 scientists in nine universities / research institutes in U.S. and Europe
  - Product in 2007
    - 1,000,000 figure / table objects, most from 2000+
    - 1,000 peer reviewed journals

# Deep Indexing Implementations

- Illustrata component search fields include
  - Object Descriptors
  - Object Geographic Terms
  - Object Statistical Terms
  - Object Subject Terms
  - Object Taxonomic Terms
  - Caption

11 results found for: (hydrothermal vent) and organism in ? CSA Illustrata: Natural Sciences - [Alert Me](#)

11 ? CSA Illustrata: Natural Sciences

Results: Article view

Published Works 11

Tables & Figures 15

All Publication Types 11

Journals 11

Peer-Reviewed Journals 9

[Mark](#) or [Clear](#) all on page | [Update Marked List](#) | [Save, Print, Email](#)

Sort by: [Most Recent First](#) Go

◀ Previous 1 Next ▶

Record #  Go

1. [Effect of ambient oxygen concentration on activities of enzymatic antioxidant defences and aerobic metabolism in the hydrothermal vent worm, Paralvinella grasslei](#)

[Marie, Benjamin](#); [Genard, Bertrand](#); [Rees, Jean-Francois](#); [Zal, Franck\\*](#)  
**Marine Biology** [Mar. Biol.]. Vol. 150, no. 2, pp. 273-284. Nov 2006.

... grasslei is a common endemic polychaete from the deep-sea hydrothermal vent communities located on the East Pacific Rise (EPR). These organisms colonise a large range of microhabitats around active sites where physico-chemical conditions are thought ...



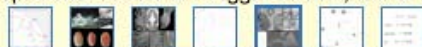
[View Record](#) | [FindText](#)

Database: CSA Illustrata: Natural Sciences  
Descriptors: [Gills](#) | [Metabolism](#) | [Antioxidants](#) | [Hydrothermal springs](#) | [Endemic species](#) | [Marine invertebrates](#) | [More...](#)

2. [Morphology, reproductive biology and genetic structure of the whale-fall and hydrothermal vent specialist, Bathykurila guaymasensis \(Annelida: Polynoidae\)](#)

[Glover, Adrian G](#); [Goetze, Erica](#); [Dahlgren, Thomas G](#); [Smith, Craig R](#)  
**Marine Ecology** [Mar. Ecol.]. Vol. 26, no. 3-4, pp. 223-234. Sep 2005.

... Polynoidae), a species recorded from both whale-falls and hydrothermal vents. The habitat of B. guaymasensis is quite different from other vent polynoids, being apparently a specialist feeder on Beggiatoa mat, rather than living commensally ...



[View Record](#) | [FindText](#)

Database: CSA Illustrata: Natural Sciences  
Descriptors: [Hydrothermal vents](#) | [Basins](#) | [Genetic structure](#) | [Haplotypes](#) | [Gametogenesis](#) | [Cytochrome-c oxidase](#) | [More...](#)

3. [An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent](#)

[Beatty, JThomas](#); [Overmann, Joerg](#); [Lince, Michael T](#); [Manske, Ann K](#); [Lang, Andrew S](#); [Blankenship, Robert E](#); [Van Dover, Cindy L](#); [Martinson, Tracey A](#); [Plumley, FGerald](#)  
**Proceedings of the National Academy of Sciences, USA** [Proc. Natl. Acad. Sci. USA]. Vol. 102, no. 26, pp. 9306-9310. Jun 2005.

... of a previously unknown green sulfur bacterial species from a deep-sea hydrothermal

Database: CSA Illustrata: Natural Sciences  
Descriptors: [Sulphur](#) | [Hydrothermal springs](#) | [Deep water](#) | [Photosynthesis](#) | [Organic carbon](#) | [Light intensity](#) | [More...](#)

Done



Database CSA Illustrata: Natural Sciences

Title **Morphology, reproductive biology and genetic structure of the whale-fall and *hydrothermal vent* specialist, *Bathykurila guaymasensis* (Annelida: Polynoidae)**

Author [Glover, Adrian G](#); [Goetze, Erica](#); [Dahlgren, Thomas G](#); [Smith, Craig R](#)

Affiliation Zoology Department, The Natural History Museum, London, UK, [a.glover@nhm.ac.uk](mailto:a.glover@nhm.ac.uk)

Source Marine Ecology [Mar. Ecol.], Vol. 26, no. 3-4, pp. 223-234, Sep 2005.

## Objects



Figure 1.



Figure 2.



Figure 3.



Figure 4.

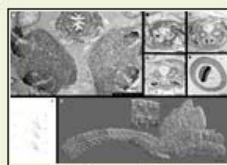


Figure 5.



Figure 6.

Haplotype diversity	
Santa Cruz	0.1121 (0.0550/0.046)
Santa Catalina	0.1054 (0.0527/0.047)
Total	0.1087 (0.0540/0.029)

Table 1.

Enhanced abstract

[Contact publisher for rights.](#)

**Abstract** We examined the reproductive biology and genetic structure of the polychaete *Bathykurila guaymasensis* (Annelida: Polynoidae), a species recorded from both whale-falls and *hydrothermal vents*. The habitat of *B. guaymasensis* is quite different from other *vent* polynoids, being apparently a specialist feeder on *Beggiatoa* mat, rather than living commensally with mussels or on the sides of hot *vent* chimneys. Specimens were collected from whale-fall sites in the Santa Cruz and Santa Catalina Basins, Southern California at depths of 1200-1600 m, and investigated using a combined morphological and molecular approach. The species exhibits marked sexual dimorphism, asynchronous gametogenesis, evidence for internal fertilization and lecithotrophic larval development - all characters shared with other *hydrothermal vent* polynoids. Two distinct, sympatrically distributed haplotype groups of *B. guaymasensis* were identified using mitochondrial cytochrome oxidase 1 gene sequences, indicating that there may be two species within the nominal designation *B. guaymasensis*. Broad sharing of haplotypes between the two whale-fall sites suggests high dispersal rates among basins along the California margin.

## Descriptors

**Article Subject Terms:**  *Hydrothermal vents*  Basins  Genetic structure  Haplotypes  Gametogenesis  Cytochrome-c oxidase  Fertilization  Mitochondria  *Hydrothermal springs*  Population genetics  Reproduction  Zoobenthos  Sexual dimorphism  Marine invertebrates  Deep water  Animal physiology  Animal morphology  Larval development

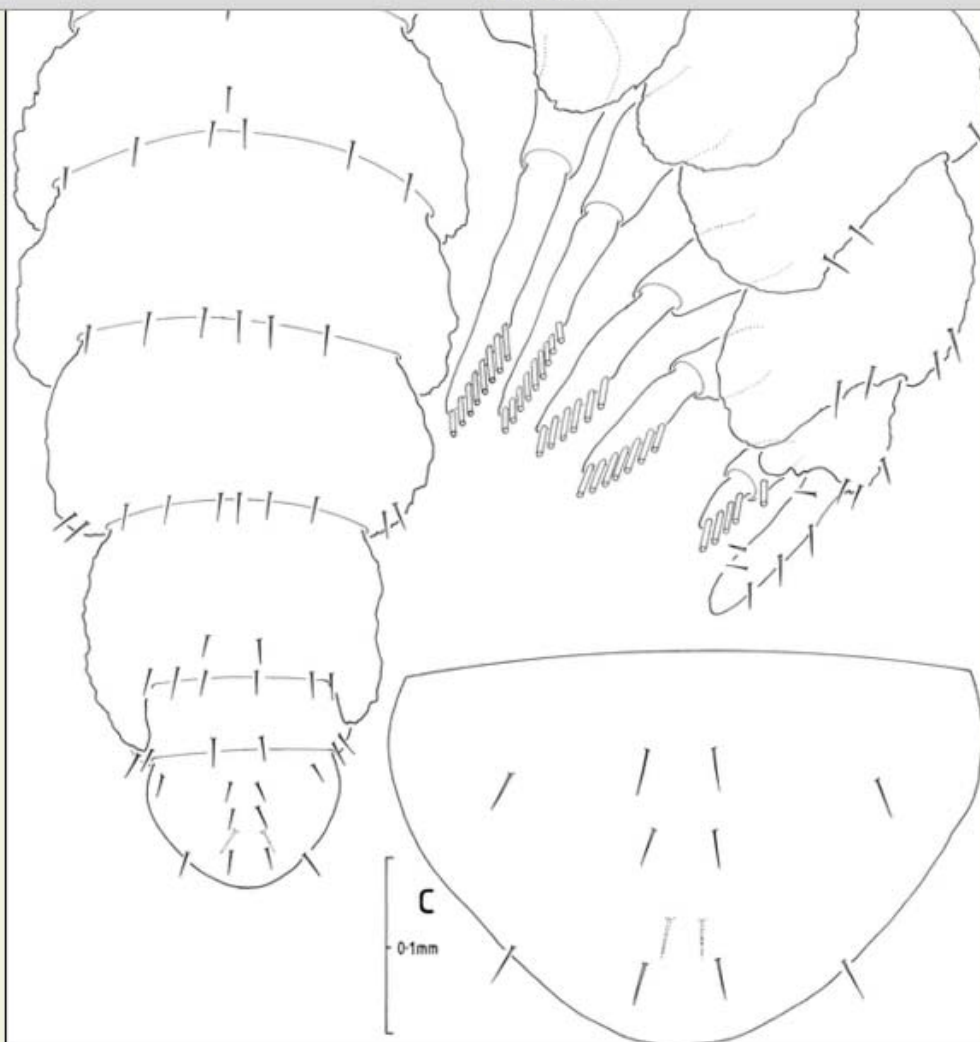
**Article Taxonomic Terms:**  Polynoidae  Annelida  *Beggiatoa*  *Bathykurila guaymasensis*

**Article Geographic Terms:**  USA, California  USA, California, Santa Cruz  INE, USA, California, Santa Catalina Basin

## Object Descriptors

**Object Subject Terms:**  Acicular chaetae  Cytochrome oxidase 1 haplotypes  Developing sperm  Enlarged nephridial

Done



Object  
record - pt. 1

Jeng, M., Clark, P. F., & Ng, P. K. (2004). The First Zoea, Megalopa, And First Crab Stage Of The Hydrothermal Vent Crab *Xenograpsus Testudinatus* (Decapoda: Brachyura: Grapsoidea) And The Systematic Implications For The Varunidae [Figure 11]. *Journal of Crustacean Biology*, 24, 188-212. Publisher: Allen Press, Inc.

[© Get Permissions](#) [Copyright Clearance Center](#)

Done

View Record

Map labels: Whim Creek, KCD-22 (2155 ± 10 Ma), RHDH2A (2146 ± 10 Ma), Marble Bar, ABOP-6 (2432 ± 14 Ma, 2179 ± 7 Ma), BD2 (2175 ± 10 Ma), Nullagine, DDH 186 (2416 ± 12 Ma), Wittenoom, WRL-1 (2194 ± 7 Ma), FVG-1 (2153 ± 10 Ma), Tom Price (2399 ± 6 Ma, 2216 ± 13 Ma), Paraburdoo, Newman (2175 ± 35 Ma).

Legend:

- Upper Wyloo Gp and overlying successions
- Lower Wyloo Group
- Turee Creek Group
- Hammersley Group
- Fortescue Group
- Mount Bruce Supergp
- Archean granite
- Archean greenstone

Formation Group	Age (Ma)
Wittenoom Fm	2561 ± 8 Ma
Marra Mamba Iron Fm	2597 ± 5 Ma
Jaeinuh Formation	2629 ± 5 Ma
Maddina & Turnbulla	2763 ± 13 Ma
Kylena & Hardry	2775 ± 10 Ma
Mt Roe Basalt	2763 ± 13 Ma
Whim Creek Gp	2775 ± 10 Ma
De Grey Gp	
Gorge Creek Gp	
Sulphur Springs Gp	3235 ± 3 Ma
Warawoona Gp	
Coonteruhah Suc.	3515 ± 3 Ma

Age (Ma) Scatter Plot:

SW → NE

Y-axis: 2100, 2150, 2200, 2250 Ma

X-axis: Tom Price, WRL-1, ABOP-6, Newman, FVG-1, BD2, KCD-22, RHDH2A

Object record - pt. 2

Rasmussen, B., Fletcher, I. R., & Sheppard, S. (2005). Isotopic dating of the migration of a low-grade metamorphic front during orogenesis [Figure 1]. , 33, 773-776.  
 Publisher: Geological Society of America

[Get Permissions](#) [Copyright Clearance Center](#)

**Category** [Figure](#); [Graph](#); [Scatter Plot](#); [Map](#)

**Title** **Isotopic dating of the migration of a low-grade metamorphic front during orogenesis**

**Author** [Rasmussen, Birger](#); [Fletcher, Ian R.](#); [Sheppard, Stephen](#)

**Source** Geology, Vol. 33, No. 10, pp. 773-776. Oct 2005.

---

**Object Descriptors**

**Object Subject Terms:**  Gascoyne Complex  Monazite dates  U-Pb zircon dates  Xenotime dates

**Object Geographic Terms:**  Australia, Western Australia, Pilbara Craton

Done

Advanced Search Results

15 results found for: (hydrothermal vent) and organism in ? CSA Illustrata: Natural Sciences [Alert Me](#)

Published Works [11](#) **Tables & Figures 15**

**Object search: results**

All 15 **Figures 15** [Enhance research with tables and figures. Take our survey.](#)

[Mark](#) or [Clear](#) all on page | [Update Marked List](#) | [Save, Print, Email](#) Sort by: **Most Recent First**

◀ Previous 1 2 Next ▶ Record #

1. [Figure 5. Reproductive features of Bathykurila guaymasensis . A: Transverse ...](#)



ORIGINAL ARTICLE: Morphology, reproductive biology and genetic structure of the whale-fall and **hydrothermal vent** specialist, *Bathykurila guaymasensis* (Annelida: Polynoidae)  
[Glover, Adrian G.](#); [Goetze, Erica](#); [Dahlgren, Thomas G.](#); [Smith, Craig R.](#)  
**Marine Ecology**, Vol. 26, No. 3-4, pp. 223-234. Sep 2005.  
[View Record](#) | [View Abstract](#) | [Find@Text](#)

Database: CSA Illustrata: Natural Sciences  
 Object Descriptors: [Developing sperm](#) | [Nephridial papilla](#) | [Ovaries](#) | [Oviduct-spermatheca](#) | [Reproductive features](#) | [Sex](#) | [More...](#)

2. [Figure 10. Xenograpsus testudinatus Ng, Huang, and Ho, 2000 , megalopa, a. chela, ...](#)



The First Zoea, Megalopa, And First Crab Stage Of The **Hydrothermal Vent** Crab *Xenograpsus Testudinatus* (Decapoda: Brachyura: Grapsoidea) And The Systematic Implications For The Varunidae  
[Jeng, M.](#); [Clark, P.F.](#); [Ng, P.K.L.](#)  
**Journal of Crustacean Biology**, Vol. 24, No. 1, pp. 188-212. 2004.  
[View Record](#) | [View Abstract](#) | [Full-Text Linking](#) | [Find@Text](#)

Database: CSA Illustrata: Natural Sciences  
 Object Descriptors: [Chela](#) | [Walking leg](#) | [Xenograpsus testudinatus](#)

3. [Figure 11. Xenograpsus testudinatus Ng, Huang, and Ho, 2000 , megalopa, abdomen, ...](#)



Database: CSA Illustrata: Natural Sciences  
 Object Descriptors: [Abdomen](#) | [View of telson](#) | [Xenograpsus testudinatus](#)

Done



# Deep Indexing Implementations

- TableSeer (CiteSeer) - 2006+
- Addressing:
  - Table search engine
  - Table metadata specification
  - Table detection / metadata extractor
    - Untagged documents
  - Table results ranking algorithm
- Not focused on interface
  - But see <http://chemxseer.ist.psu.edu/>

# Deep Indexing Implementations

- BioText Search Engine - 2007+
  - 80,000 objects
  - 20,000 articles
  - 150 open access journals from PubMed Central
  - Growth in collection size & function

BioText Search Engine

[Home](#) | [About BioText](#) | [Contact Us](#)

**BioText** SEARCH ENGINE

Search:

Search Over:  Full Text & Abstracts  Figure Captions (List)  Figure Captions (Grid)  Tables Sort By: Relevance

Results 1-7 of 7  ABSTRACTS  FULL-TEXT EXCERPTS  FIGURES

---

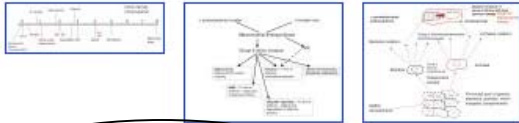
**The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate**  
 Koonin, E. (2006) *Biology Direct*.

**ABSTRACT**  
 Ever since the discovery of 'genes in pieces' and mRNA splicing in eukaryotes, origin and evolution of spliceosomal introns have been considered within the conceptual framework of the 'introns early' versus 'introns late' debate. The 'introns early' hypothesis, which is closely linked to the so-called exon theory of gene evolution, posits that protein-coding genes were interrupted by numerous introns even at the earliest stages of life's evolution and that introns played a major role in the origin of proteins by facilitating recombination of sequences coding for small protein/peptide modules. Under this scenario, the absence of spliceosomal introns in prokaryotes... [Show Full Abstract](#)

**FULL-TEXT EXCERPTS**  
 ...for bacteria than it is for organelles [57]. So what happened at the onset of eukaryotic evolution? The scenarios of **eukaryogenesis** favored by different researchers differ dramatically, and this is not the place to review in any detail the pros and cons of each of these scenarios [58]. In any case, there is no doubt that the symbiosis between an  $\alpha$ -proteobacterium, the ancestor of the mitochondria, and a somewhat mysterious host, the ancestor of the eukaryotic... [Show Full Excerpts](#)

VIEW FULL ARTICLE: [HTML](#) | [PDF](#)

FIGURES FROM ARTICLE:



[View all 3 figure\(s\) with captions](#)


Results: Article view

---

**The ancient Virus World and evolution of cells**  
 Koonin, E., Senkevich, T., Dolja, V. (2006) *Biology Direct*.

**ABSTRACT**  
 Recent advances in genomics of viruses and cellular life forms have greatly stimulated interest in the origins and evolution of viruses and, for the first time, offer an opportunity for a data-driven exploration of the deepest roots of viruses. Here we briefly review the current views of virus evolution and propose a new, coherent scenario that appears to be best compatible with comparative-genomic data and is naturally linked to models of cellular evolution that, from independent considerations, seem to be the most parsimonious among the existing ones. Several genes coding for key proteins involved in

FIGURES FROM ARTICLE:



[View all 3 figure\(s\) with captions](#)

Done

BioText Search Engine

[Home](#) | [About BioText](#) | [Contact Us](#)

**BioText** SEARCH ENGINE

Search

Search Over:  Full Text & Abstracts  Figure Captions (List)  Figure Captions (Grid)  Tables Sort By: Relevance

Results 1-1 of 1

Enhanced abstract

**The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?**

Koonin, E. (2006) *Biology Direct*.

**ABSTRACT**

Ever since the discovery of 'genes in pieces' and mRNA splicing in eukaryotes, origin and evolution of spliceosomal introns have been considered within the conceptual framework of the 'introns early' versus 'introns late' debate. The 'introns early' hypothesis, which is closely linked to the so-called exon theory of gene evolution, posits that protein-coding genes were interrupted by numerous introns even at the earliest stages of life's evolution and that introns played a major role in the origin of proteins by facilitating recombination of sequences coding for small protein/peptide modules. Under this scenario, the absence of spliceosomal introns in prokaryotes... [Show Full Abstract](#)

VIEW FULL ARTICLE: [HTML](#) | [PDF](#)

**Figures From Article (3)**

**FIGURE CAPTION**

Figure 1. Intron density in the genes of some modern eukaryotes and the reconstructed common ancestor of plants and animal (possible LECA under unicont-bicont phylogeny). The marks on the line show approximate intron density values (mean number of introns per gene); the... [Show Full Caption](#)

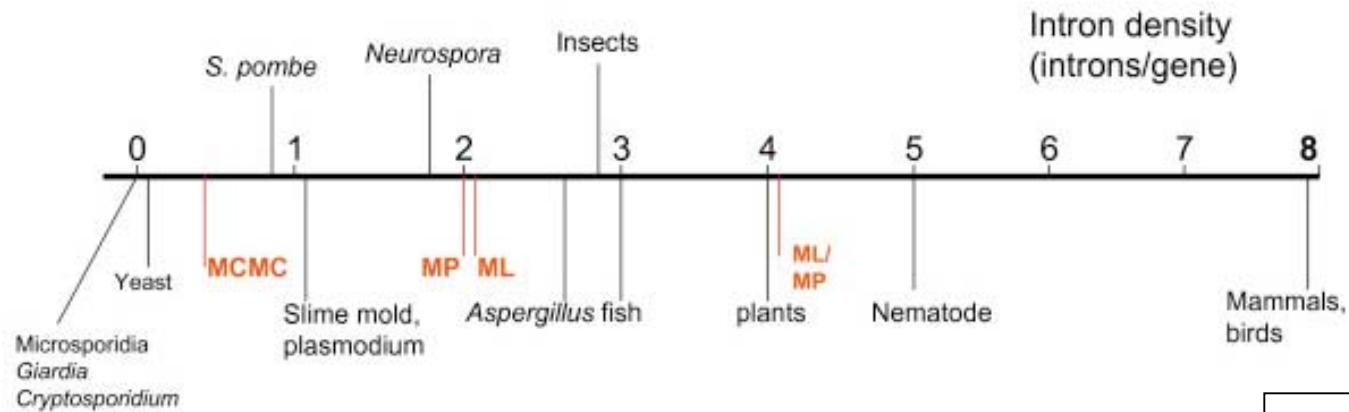
**FIGURE CAPTION**

Figure 2. The proposed chain of causes and events in eukaryogenesis – the pivotal roles of mitochondrial endosymbiosis and intron invasion. Arrows indicate proposed causal relationships (selective forces).

**FIGURE CAPTION**

Figure 3. A brief early history of spliceosomal introns. The scheme shows the...

Done



Object record

Figure 1. Intron density in the genes of some modern eukaryotes and the reconstructed common ancestor of plants and animal (possible LECA under unicont-bicont phylogeny). The marks on the line show approximate intron density values (mean number of introns per gene); the data is from [99]. The reconstructed values for the ancestral genome (in red) are denoted after the reconstruction method: MCMC, Markov Chain Monte Carlo[49]; MP, maximum parsimony[23]; ML, maximum likelihood[41]; ML/MP, hybrid maximum likelihood/parsimony method[51].

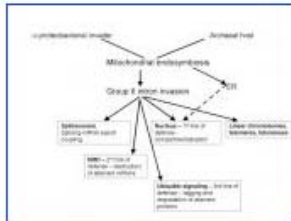
eukaryogenesis

Search

Search Over:  Full Text & Abstracts  Figure Captions (List)  Figure Captions (Grid)  Tables Sort By: Relevance

Results 1-3 of 3

Object search: results as list



[View Fig. & Caption](#)

### The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate

Koonin, E. (2006) *Biology Direct*.

#### FIGURE CAPTION

Figure 2. The proposed chain of causes and events in **eukaryogenesis** – the pivotal roles of mitochondrial endosymbiosis and intron invasion. Arrows indicate proposed causal relationships (selective forces).

VIEW FULL ARTICLE: [HTML](#) | [PDF](#) ([View all figures and tables from this article](#))



[View Fig. & Caption](#)

### The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate

Koonin, E. (2006) *Biology Direct*.

#### FIGURE CAPTION

Figure 1. Intron density in the genes of some modern eukaryotes and the reconstructed common ancestor of plants and animal (possible LECA under unicont-bicont phylogeny). The marks on the line show approximate intron density values (mean number of introns per gene); the data is from [99]. The reconstructed values for the ancestral genome (in red) are denoted... [Show Full Caption](#)

VIEW FULL ARTICLE: [HTML](#) | [PDF](#) ([View all figures and tables from this article](#))



### The origin of introns and their role in eukaryogenesis: a compromise

Done

virus

Search

Search Over:  Full Text & Abstracts  Figure Captions (List)  Figure Captions (Grid)  Tables Sort By: Relevance

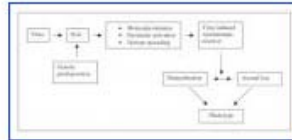
Results 1-20 of 2834 << Prev 1 2 3 4 5 Next >>

Object search: results as grid



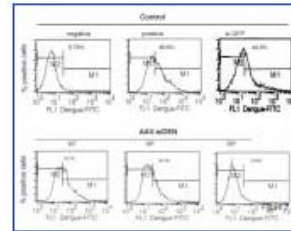
[View Fig. & Caption](#) | [View All Figs.](#)

Figure 1. Infectious scenarios and treatment window for patients coinfected with hepatitis B and...



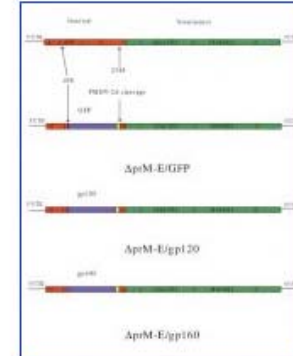
[View Fig. & Caption](#) | [View All Figs.](#)

Figure 1. Proposed scheme for virus-mediated autoimmunity in multiple...



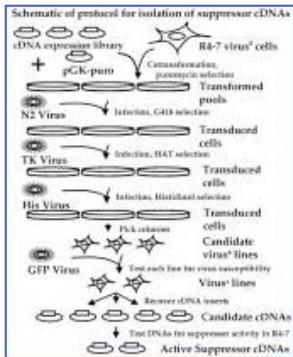
[View Fig. & Caption](#) | [View All Figs.](#)

Figure 2. AAVsIDEN expression decreases DEN-2 virus infection in cultured Vero cells. Cells were...

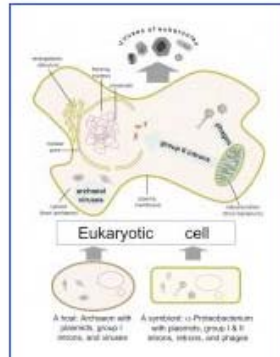


[View Fig. & Caption](#) | [View All Figs.](#)

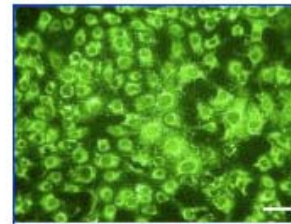
Figure 1. Construction of wild type dengue virus and dengue virus replicon vectors used in these...



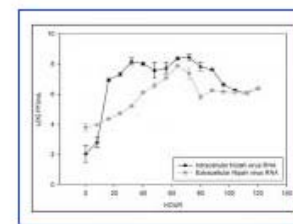
[View Fig. & Caption](#) | [View All Figs.](#)



[View Fig. & Caption](#) | [View All Figs.](#)



[View Fig. & Caption](#) | [View All Figs.](#)



[View Fig. & Caption](#) | [View All Figs.](#)

Done

virus

Search

Search Over:  Full Text & Abstracts  Figure Captions (List)  Figure Captions (Grid)  Tables Sort By: Relevance

Results 1-20 of 1031 << Prev 1 2 3 4 5 Next >>

TABLES EXPANDED

**Host-virus interaction: a new role for microRNAs**

Scaria, V., Hariharan, M., Maiti, S., Pillai, B., Brahmachari, S. (2006) *Retrovirology*.

TABLE CAPTION

Table 1. List of **virus**-encoded microRNAs

Source Virus	Virus Type	Number of microRNAs	References
Epstein Barr <b>virus</b>	Herpesvirus	32	[42]
Kaposi sarcoma-associated herpesvirus	Herpesvirus	17	[43]
Mouse gammaherpesvirus	Herpesvirus	10	[42]
Human cytomegalovirus	Herpesvirus	14	[42]
Herpes Simplex-1	Herpesvirus	1	[44]
Rhesus lymphocryptovirus	Herpesvirus	22	[45]
Simian <b>virus</b> 40	Papovavirus	2	[41]

[Show Full Table \(9 rows\)](#)

VIEW FULL ARTICLE: [HTML](#) | [PDF](#) [\(view all figures and tables from this article\)](#)

Object search: table results

**The ancient Virus World and evolution of cells**

Koonin, E., Senkevich, T., Dolja, V. (2006) *Biology Direct*.

TABLE CAPTION

Table 4. Major concepts in **virus** evolution

Concept	Principal message	References	Brief critique/comment
Cell degeneration	Viruses, at least complex ones, evolved as a		This route of <b>virus</b> evolution appears to be inconsistent with

Stopped

# Interface Comparison

- BioText and Illustrata
- “Enhanced abstract” view of same article

Search

Search Over:  Full Text & Abstracts  Figure Captions (List)  Figure Captions (Grid)  Tables Sort By: Relevance

Results 1-1 of 1

**The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?**Koonin, E. (2006) *Biology Direct*.**ABSTRACT**

Ever since the discovery of 'genes in pieces' and mRNA splicing in eukaryotes, origin and evolution of spliceosomal introns have been considered within the conceptual framework of the 'introns early' versus 'introns late' debate. The 'introns early' hypothesis, which is closely linked to the so-called exon theory of gene evolution, posits that protein-coding genes were interrupted by numerous introns even at the earliest stages of life's evolution and that introns played a major role in the origin of proteins by facilitating recombination of sequences coding for small protein/peptide modules. Under this scenario, the absence of spliceosomal introns in prokaryotes... [Show Full Abstract](#)

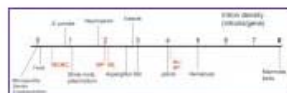
VIEW FULL ARTICLE: [HTML](#) | [PDF](#)**Figures From Article (3)****FIGURE CAPTION**

Figure 1. Intron density in the genes of some modern eukaryotes and the reconstructed common ancestor of plants and animal (possible LECA under unicont-bicont phylogeny). The marks on the line show approximate intron density values (mean number of introns per gene); the... [Show Full Caption](#)

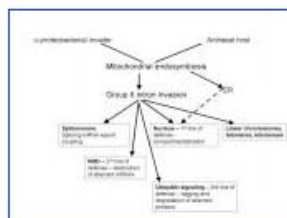
**FIGURE CAPTION**


Figure 2. The proposed chain of causes and events in eukaryogenesis – the pivotal roles of mitochondrial endosymbiosis and intron invasion. Arrows indicate proposed causal relationships (selective forces).

**FIGURE CAPTION**

Figure 3. A brief early history of spliceosomal introns. The scheme shows the

Done

View Record



What you need to know about 'Ask Your Doctor' ads

[The University of Tennessee](#)

[Please log in to My Research](#)

0 Marked Records | [Search History](#)

[Logout](#) [Quick Search](#) [Advanced Search](#) [Search Tools](#)

[Return to View Record](#) [Help & Support](#)

**Record View**

Abstract  Mark This Record | [Update Marked List](#) | [Save, Print, Email](#)

**FindText**

**Database** CSA Illustrata: Natural Sciences

**Title** **The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?**

**Author** [Koonin, Eugene V](#)

**Affiliation** National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

**Source** Biology Direct [Biol. Direct]. Vol. 1, [np]. 2006.

**Objects**

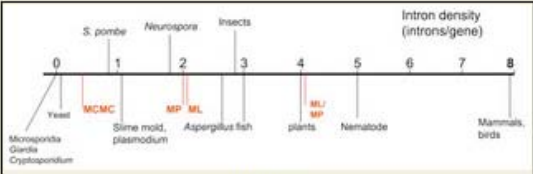


Figure 1.

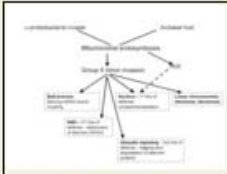


Figure 2.

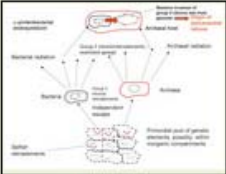


Figure 3.

Contact the publisher for rights.

**Abstract** Background Ever since the discovery of 'genes in pieces' and mRNA splicing in eukaryotes, origin and evolution of spliceosomal introns have been considered within the conceptual framework of the 'introns early' versus 'introns late' debate. The 'introns early' hypothesis, which is closely linked to the so-called exon theory of gene evolution, posits that protein-coding genes were interrupted by numerous introns even at the earliest stages of life's evolution and that introns played a major role in the origin of proteins by facilitating recombination of sequences coding for small protein/peptide modules. Under this scenario, the absence of spliceosomal introns in prokaryotes is considered to be a result of "genome streamlining". The 'introns late' hypothesis counters that spliceosomal introns emerged only in eukaryotes, and moreover, have been inserted into protein-coding genes continuously throughout the

Done

# Opportunities and dangers

- End-user evaluation (DeLiver; Illustrata):
  - Higher precisions searching
  - F & T help relevance judgments
  - Direct access to high-quality artifact
    - Without getting paper
  - F & T use in teaching / presentation

# Opportunities and dangers

- End-user evaluation (DeLiver; Illustrata):
  - Disaggregated components are re-used in new text and performative documents
    - Papers, proposals, presentations
  - Support comparisons of many kinds
    - Sanity checks; find useful data (tables)
  - Creating data compilations
  - Finding data is cheaper than generating it

# Opportunities and dangers

- However
  - Interface complexity
    - Filters (e.g., categories; dates) important
    - Illustrata vs. BioText
  - Broader communications impacts
    - Effect of decontextualization of components
    - Citing components? How and what creator benefit?
    - Preparing components that can stand alone
  - Information professionals
    - Promoting and supporting new discovery systems
    - Understanding disciplinary differences

# Opportunities and dangers

- However
  - Incommensurate implementations
  - Lack of standards
  - Lack of openness
    - Proprietary system has greatest functionality

# Issues for research and development

- The road to universal discovery
  - Metadata standards for independently findable, first-class objects / components
    - Tables, figures, maps, photographs, graphs now
    - What about
      - Datasets
      - Workflows
      - Simulations
      - Digitized specimens
      - Others?
  - Integration of component / object search and retrieval into next generation discovery tools

# Issues for research and development

- Open extraction and indexing systems
  - Precision / recall test for technical evaluation
  - Need comparative, contextualized user studies (e.g., compare BioText to Illustrata)

System	End Users
DeLiver	108 + 350 surveys
Illustrata	60
BioText	8
TableSeer	0

# Issues for research and development

- Between-discipline studies
  - Patterns of component discovery and use
    - Object types
    - Discovery and use patterns
- Improved contextualization
  - DeLiver evaluation as model
  - Component systems as part of an ecology
    - BioText and Illustrata

# Issues for research and development

- It's a question of end-user benefits & costs:
  - Does focused searching on components yield improved results (better, faster)?
  - Does component availability transform scientific practice in positive ways?

OR

- Are these additional separate systems to search?
- Does component searching create larger results sets to evaluate?

# Issues for research and development

- Practice transformations: writing & publishing to ease components use
  - Create objects as if they will stand alone
  - Publishing processes support automatic extraction and downstream processing
  - Use persistent object identifiers to cross-link
    - Articles
    - Components
    - Datasets

# Sources

## DeLlver

Bishop, A. P. (1999). Document structure and digital libraries: How researchers mobilize information in journal articles. *Information Processing and Management*, 35(3), 255-279.

Bishop, A. P., Neumann, L. J., Star, S. L., Merkel, C., Ignacio, E., & Sandusky, R. J. (2000). Digital libraries: Situating use in changing information infrastructure. *Journal of the American Society for Information Science*, 51(4), 394-413.

## Illustrata

Sandusky, R. J., & Tenopir, C. (in press). Finding and using journal-article components: Impacts of disaggregation on teaching and research practice. *Journal of the American Society for Information Science and Technology*.

Sandusky, R. J., Tenopir, C., & Casado, M. M. (2007). Figure and table retrieval from scholarly journal articles: User needs for teaching and research. Proceedings 70th Annual Meeting of the American Society for Information Science and Technology, Milwaukee, WI.

Sandusky, R. J., Tenopir, C., & Casado, M. M. (2007). Uses of figures and tables from scholarly journal articles in teaching and research. Proceedings 70th Annual Meeting of the American Society for Information Science and Technology, Milwaukee, WI.

# Sources

## TableSeer (CiteSeer)

Liu, Y., Bai, K., Mitra, P., & Giles, C. L. (2007). TableSeer: Automatic table metadata extraction and searching in digital libraries. *Proceedings of the 2007 Conference on Digital Libraries*, 91-100.

## BioText Search Engine

Hearst, M.A., Divoli, A., Ye, J., & Wooldridge, M.A. (2007). Exploring the efficacy of caption search for bioscience journal search interfaces. *BioNLP Workshop at ACL 2007*, Prague, Czech Republic, June 2007.

Hearst, M.A., Divoli, A., Guturu, H., Ksikes, A., Nakov, P., Wooldridge, M.A. and Ye, J. (2007) "BioText Search Engine: beyond abstract search", *Bioinformatics*, 23: 2196-2197.

# Contact

Dr. Robert J. Sandusky

Assistant University Librarian for Information Technology

Clinical Associate Professor

Richard J. Daley Library

University of Illinois at Chicago

[sandusky@uic.edu](mailto:sandusky@uic.edu)