

# JAbbr

How to decipher journal abbreviations  
(using metadata extracted from the Library Catalog)

by Keith Jenkins

Presented to the Metadata Working Group of the Cornell University Library, 2007-10-19

# The Problem:

Citations often use abbreviations instead of the full journal title.

# Printed bibliographies

Publishers wanted to  
reduce publication  
costs

Fewer letters =  
fewer pages to print  
less paper

**Rieneck**, — Ueber die Schichtung des Forellenkeims. Arch. Mikr. Anat., 1869, **5**, 356-366. fig. 1869.1

**Riess**, *Johann Albin* [1853 —] Der Bau der Kiemenblätter bei den Knochenfischen. Inaug. Dissert. Bonn, 1881. 35 p. 3 pls. 8°. — Arch. Naturgesch., 1881, **65**, Jahrg., 518-550. 3 pls. 1881.1

— Ueber einige fossile Chimaeriden-Reste im Münchener paläontologischen Museum. Paläontographica, 1887, **34**, 1-27. 3 pls. 1887.1

**Rietstap**, *Gerard H. A. N.* De haring en de haringvisserij. Tijdschr. Nederl. Maatsch., 1864, **27**, 312-338. 1864.1

— De zeevisserij en de makreel. Tijdschr. Nederl. Maatsch., 1870, **33** (3. ser. **11**), 122-150. 1870.1

**Riggio**, *Giuseppe*. Corrispondenze scientifiche moderne degli animali figurati nel Panphyton siculum del Cupani (Pesci) Nat. Sicil., 1882, **11**, 160-163. 1882.1

Refers to the natural history work of Francisco Cupani published at Palermo in 1713.

— Cattura di *Carcharodon rondesteti* Mull. Henl. nelle acque di capo Gallo e di isola delle Femine. Nat. Sicil., 1894, **13**, 130-133. 1894.1

— Sopra un caso di notevole ramificazione dei ciechi pilorici di *Centrolophus pompilus* Cuv. e Val. Nat. Sicil., 1894, **13**, 206-211. 3 figs. 1894.2

Pichthyologie, la culture du poisson, etc., Paris, 1869. 8°. 1869.1

— La mer et les poissons. Bull. Soc. Acad. Var., 1870, n. s. **3**, 237-282. 1870.1

— Les causes de la cherté du poisson. Bull. Soc. Acclim. Paris, 1875, 3. sér. **2**, 652-667. 1875.1

— Om aarsagerne til fiskemangelen. Nord. Tidsskr. Fiskeri, 1876, n. s. 3. Jahrg., 80-96. 1876.1

**Ringel**, *Ernst*. Zwei neue Goldfisch-Varietäten. Blätt. Aquar.-Terrar. Freunde, 1901, 12. Jahrg., 91-92. 1901.1

— Unsere Lebendgebärenden. Wochenschr. Aquar.-Terrar. Kunde, 1911, 8. Jahrg., 248-250; 258-260; 275-277; 290-293. 10 figs. 1911.1

— Unsere Haplochilus-Arten, ihre Haltung, Pflege und Zucht. Wochenschr. Aquar.-Terrar. Kunde, 1912, 9. Jahrg., 267-269; 279-281; 293-294. 19 figs. 1912.1

**Ringer**, *Sydney*. The influence of saline media on fishes. Journ. Physiol., London, 1883, **4** (Suppl.), vi-viii. 1883.1

— Concerning the influence of saline media on fish. Journ. Physiol., London, 1884, **5**, 98-115. 1884.1

**Ringer**, *W. E.*, & **Herwerden**, *M. A.* See **Herwerden & Ringer**.

# Legacy databases

J-ABN-AND-SOC-PSYCHOL

J-ABN-PSYCHOL-AND-SOC-PSYCHOL

J-AM-MED-ASS

J-AM-MED-ASS-CHICAGO

J-AM-MED-ASSOC

J-AMER-FOLK-LORE

J-AMER-FOLKLORE

J-AMER-J-OF-PHYSIOL

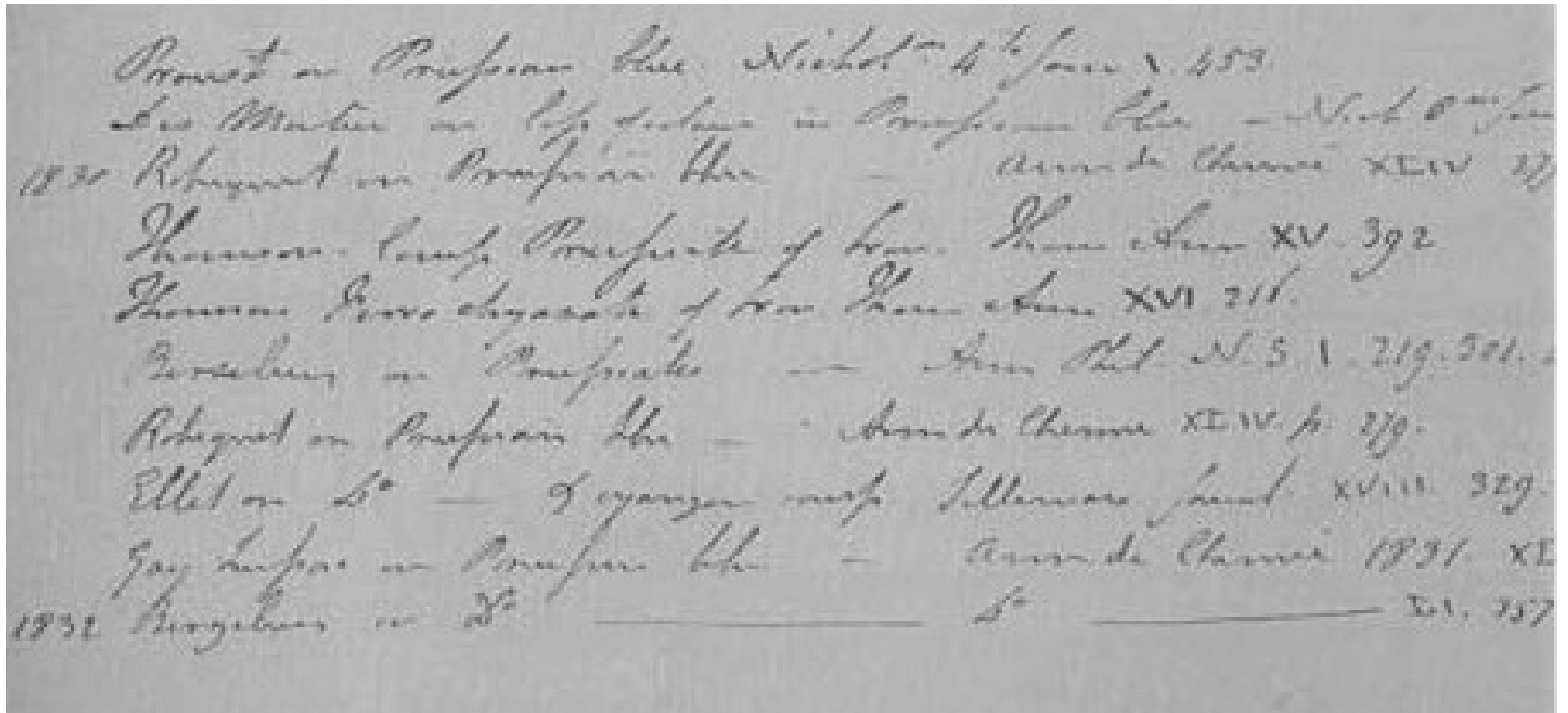
J-AMER-MED-ASS

Database providers wanted to reduce storage and processing costs

Fewer letters =

- less disk space
- less data to process
- faster retrieval

# Handwritten notes



Provest on Prussian blue. Nichol. 4<sup>th</sup> Jan 1859.  
Les Maitre on top of blue in Prussian blue. Nichol 8<sup>th</sup> Jan  
1830 Reiquart on Prussian blue. Annals Chem. XLIV 27.  
Thomson. Comp. Properties of Iron. Phil. Mag. XV. 392.  
Thomson. Pure character of Iron. Phil. Mag. XVI 216.  
Boscovich on Prussian blue. Ann. Phil. N.S. I. 219. 211.  
Reiquart on Prussian blue. Annals Chem. XLIV. p. 279.  
Eller on S<sup>o</sup>. of oxygen. Phil. Mag. XVIII. 329.  
Gay Lussac on Prussian blue. Annals Chem. 1831. XI  
1832 Berzelius on S<sup>o</sup>. S<sup>o</sup>. 21. 257

Researchers want to save time,  
and try fit everything onto a single scrap of paper

# Why abbreviate?

- To save space
  - in printed bibliographies
  - in databases
  - in hand-written notes
- To save time
  - in hand-written notes

OK, so I have my abbreviation...  
Now what?

The meaning might be obvious

J. Am. Chem. Soc. =

Journal of the American Chemical Society

(which is also commonly abbreviated as JACS)

OK, so I have my abbreviation...  
Now what?

The meaning might be obvious

Proc. Am. Soc. Psych. Res. =

Proceedings of the American Society of  
Psychological Research

OK, so I have my abbreviation...  
Now what?

The meaning might be obvious (or not)

Proc. Am. Soc. Psych. Res. =

Proceedings of the American Society for  
Psychical Research

OK, so I have my abbreviation...  
Now what?

Proc. Soc. Natl. Hist. =

Proceedings of the Boston? Society of Natural History

Proceedings of the Portland? Society of Natural History

Proceedings of the San Diego? Society of Natural History

(The abbreviation may have been transcribed incompletely.)

# Existing tools for abbreviations

Many reference works  
can be found in the catalog:

subject = Periodicals--Abbreviations of titles

# Existing tools for abbreviations

"All that JAS" (Journal Abbreviation Sources)  
<http://www2.iastate.edu/~cyberstacks/JAS.htm>

# Limitations of Abbreviation Lists

- Often don't include more obscure titles
- Often don't include proceedings
- May not include older forms (title changes)
- Often segregated by subject
- Only include "official" abbreviations

# A different approach

Given an abbreviation,

go through a list of ALL journals,

selecting those titles

that MIGHT match the abbreviation.

A list of ALL journals?

Or, at least as many as we can get.

# A large list of titles...

- There are over 300,000 serials records in the library catalog.
  - journals
  - proceedings
  - other serial works whose titles might get abbreviated

# The Strategy

- Extract serial title information from the catalog
- Create a web interface:
  - The user enters their cryptic abbreviation
  - Apply fuzzy matching against the titles
  - Return list of possible matches
  - The user selects the best match
  - View the full catalog record for the title

# Extracting the titles

- MARC record format stores title information in many fields:
  - 245 = Title statement (transcribed from item)
  - 210 = Abbreviated title
  - 222 = Key title
  - 240 = Uniform title
  - 242 = Translation of title by cataloging agency
  - 243 = Collective uniform title
  - 246 = Varying form of title
  - 247 = Former title

# Extracting the titles

- Export MARC records from the catalog
- Convert MARC to MARCXML
- Convert MARCXML to an efficient format for searching

# Exporting MARC records

- Use VgerSelect to export every serial record in the catalog
  - more than 300,000 records
  - about 3 hours to run
  - 400MB
- Can periodically run an update to get new records
  - one year ~ 30,000 records, 25 minutes, 40MB

# VgerSelect

**Select Voyager records: JAbbr: bibid, status, MARC**

Select by Voyager date and general criteria

<input checked="" type="checkbox"/> Created on or after:	<input type="checkbox"/> Created on or before:
2006 AUG 25	2000 DEC 31
Apply also to: <input type="text"/>	Apply also to: <input type="text"/>

OR  AND

<input type="checkbox"/> Updated on or after:	<input type="checkbox"/> Updated on or before:
1999 JAN 31	1999 JAN 31
Apply also to: <input type="text"/>	Apply also to: <input type="text"/>

Select distinct

Table name prefix:     Create operator ID(s):     Update operator ID(s):

> Next  
< Previous  
<< First  
>> Last  
  
Process  
Exit

# VgerSelect

**Select Voyager records: JAbbr: bibid, status, MARC**

Select by record type and bibliographic level

Include records with these:

Record type:

- a Language
- c Music
- d MS Music
- e Cartographic
- f MS Cartographic
- g Projected

Bibliographic level:

- b Serial component
- c Collection
- d Subunit
- i Integrating resource
- m Monograph
- s Serial**

Exclude records with these:

Record type:

- a Language
- c Music
- d MS Music
- e Cartographic
- f MS Cartographic
- g Projected

Bibliographic level:

- a Mono component
- b Serial component
- c Collection
- d Subunit
- i Integrating resource
- m Monograph

> Next

< Previous

<< First

>> Last

Process

Exit

# VgerSelect

```
20071018 15:23:53 Job began
20071018 15:47:46 Job ended
                00:23:53 Elapsed time for job

20071018 15:23:53 Search (SQL step) began
20071018 15:23:53 Search (SQL step) ended
                00:00:00 Elapsed time for search step

20071018 15:23:53 Examination of records began
20071018 15:47:46 Examination of records ended
                00:23:53 Elapsed time for examination
                0.0480227882 Average time (seconds) to examine each record
                20.8234473133 Records examined per second
                1,249.4068387997 Records examined per minute
```

## Initial SQL statement:

```
SELECT DISTINCT T1.BIB_ID FROM CORNELLDB.BIB_MASTER T1,
CORNELLDB.BIB_TEXT T2 WHERE T1.CREATE_DATE >= TO_DATE('2006
08 25', 'YYYY MM DD') AND T2.BIB_FORMAT IN ('as', 'cs', 'ds',
'es', 'fs', 'gs', 'is', 'js', 'ks', 'ms', 'os', 'ps', 'rs',
'ts') AND T1.BIB_ID = T2.BIB_ID
```

```
0 records to be read
29840 records actually read
```

```
29840 accepted for inclusion in output file
```

```
29840 MARC-format records written to output file:
```

```
c:\jabbr\10\marc\vgersele.2007-10-18.20071018.152353.marc.mrc
```

```
29840 lines written to output file:
```

```
c:\jabbr\10\marc\vgersele.2007-10-18.20071018.152353.data.txt
```

# marc2xml.php

- Written to unpack the MARC structure and output MARCXML
- Option to only include specified fields in the XML
  - reduces file size and processing time
  - Currently using 210a, 245abnp, 246a, 247a
- 400MB MARC yields 90MB XML

# marc2xml.php

<record>

<controlfield tag="001">5852609</controlfield>

<datafield tag="210" ind1="0" ind2=" ">

<subfield code="a">Liaison - Can. Centre Occup. Health Saf.</subfield>  
</datafield>

<datafield tag="245" ind1="0" ind2="0">

<subfield code="a">Liaison :</subfield>  
<subfield code="b">a newsletter for the users of CCINFO /</subfield>  
</datafield>

<datafield tag="246" ind1="0" ind2="5">

<subfield code="a">Liaison</subfield>  
</datafield>

</record>

# xml2jabbr.php

- Normalizes the titles
- Reorganizes data into a more efficient structure for searching
  - For each title, there is a tab-delimited line:
    - Journal of American history <TAB> journal of american history <TAB> j am hist
    - Journal of applied non-classical logics <TAB> journal of applied non classical logics
- 90MB XML yields 27MB "JAbbr" data

# mergeJabbr.php

- Merges multiple JAbbr files (original + updates)
- Splits into separate files for each initial letter
  - Approx 1MB per file
  - Pre-sorted alphabetically
  - Less work for each query
  - Some titles appear in multiple files, if an alternate title starts with a different letter. For example, the "D" file has:

AETA directory<TAB>directory of the american  
educational theatre association

# service.php

- Accepts two parameters:

q = the abbreviation

max = maximum results to return

- Uses regular expressions to find matches

- Returns XML: 

```
- <list q="agr eng j" max="20">
  <q regexp="\bag\w*r[^\t]* en\w*g[^\t]* j"/>
  - <item>
    <title>Agricultural engineering journal</title>
  </item>
  - <item>
    <title>Philippine agricultural engineering journal</title>
    <altTitle>agricultural engineering journal</altTitle>
  </item>
</list>
```

# The Web Interface

- Uses javascript for dynamic response ("AJAX")
- Fallback to standard form submission, if javascript not available
- Quick response
- Alphabetical list of results; clean display
- Links to catalog records

<http://supportingcast.mannlib.cornell.edu/jabbr/>

# Integration into Voyager Catalog

- Worked with Rick Silterra
- Added "Journal Title Abbreviation" option in basic search
  - Rick wrote Javascript to re-route to JAbbr
  - Keith styled JAbbr to look like the catalog
  - Worked together to persist the existing Voyager session

<http://catalog.library.cornell.edu/>

# Some limitations

- Had to adjust the formula to handle:
  - Natl, Engr, Boln
  - Jrnl, Kgl
  - FATUS (and other initialisms)
- Still fails occasionally
- Only finds titles that Cornell owns

# Future Development?

- Incorporate titles from other sources
  - All those online lists
  - OCLC WorldCat?

# Contact

Please refer all questions and suggestions to

Keith Jenkins  
kgj2@cornell.edu