

Creating a unique title search tool using Alma Analytics

Alvin Stockdale & Mike Davidson
Librarians, National Library of
Medicine

IGeLU 2022

September 2022

Use the chat window for questions.

As time permits, questions will be addressed at the end of the meeting.



U.S. National Library of Medicine

Agenda

- Resource Description and Access (RDA) requirement
- Why existing Alma/Analytics fields don't solve this problem
- Creating a unique title search tool
- QA the results and iterate on the formula
- NERS enhancement and idea exchange
- Q&A



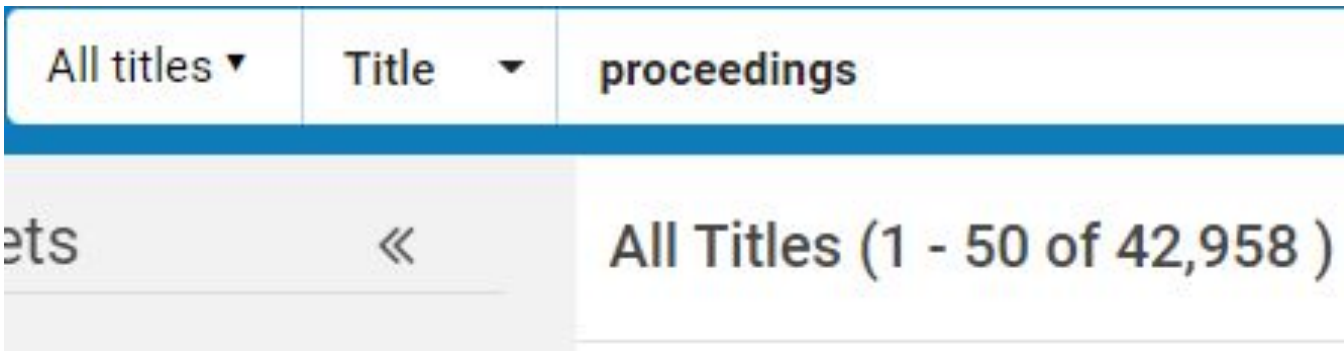
Unique Access Points

- RDA requirement
- MARC 245 field must be searched to see if the title is unique, if it is not unique, it is qualified with a 1XX or 240 field.
- The first step in creating a unique access point is searching your title to determine if it is unique.
- The MARC 245 subfields used to determine title uniqueness are 245 \$a, \$n, \$p.



Title search in Alma

- The Alma MARC 21 search index for the title field searches 38 fields and over 100 subfields.
- Difficult to search titles with one or two words.
- Basic title search of the word “proceedings” returns 43,000 titles in our database. Title = “proceedings” returns over 1,000



What does a unique title search need to do?

- Normalize the MARC 245 field and put all letters in lowercase
- Remove initial articles in all languages (A, An, The, etc.)
- Pack consecutive whitespace characters into one space
- Only return portion of the title field that determines uniqueness (245 \$a, \$n, \$p)



Unique Title Search Tool

Unique Title Search Tool

Enter your 245 \$a (or \$a \$n \$p if applicable) into the search box in **all lowercase letters** and with **no punctuation or diacritics**. **Do not include initial articles**. If your title has \$n and \$p as well, **enter those values without punctuation, diacritics, or subfield symbols**. **Do not enter \$b or \$c values** if they exist. Once you enter your value in the search box, **hit the enter key**.

Using a custom Title (Complete) field allows for the best search to see if your title is unique. For example, a search of the word **sida** brings back the following 245 \$a values.

Le SIDA
Le S.I.D.A.
SIDA
Sida
Sida

Examples below show the 245 field in a MARC record and the second line shows what you should search in this tool.

245 00 \$a Nanomedicine : \$ nanotechnology, biology, and medicine.

Search: nanomedicine

245 04 \$a The journal of politics.

Search: journal of politics

245 00 \$a Proceedings of the Institution of Mechanical Engineers. \$n Part B, \$p Journal of engineering manufacture.

Search: proceedings of the institution of mechanical engineers part b journal of engineering manufacture

245 13 \$a al-Tārīkh al-Islām

Search: tarikh alislam

245 03 \$a La Mère et l'enfant

Search: mere et lenfant

245 10 \$a Ch'ōnsadūl ūn uri yōpchip e sanda : \$b sahoejōk t'ūrauma ūi ch'i'yu rūl wihayō / \$c Chōng Hye-sin, Chin Ūn-yōng chiūm.

Search: chonsadul un uri yopchip e sanda

Page 1

* Title (Complete) regex

OK

Reset ▾



What results should be returned?

- Cataloger has a new book with the title of SIDA. They enter sida in the tool and get the following titles in the search results.
- SIDA
- Sida
- Le sida
- El SIDA
- SIDA!?
- Sídà
- Le SIDA
- El sida
- La SIDA
- Le SIDA
- Le S.I.D.A.
- Du sida



Search results page

245 \$a \$n \$p title search

total 141

NLM Unique ID (035 \$9)	MMS id	Title (Complete)	Author	Uniform Title	Title (Complete) regex
100898526	998985263406676	Le sida : guide du praticien : diagnostic, traitement, prise en charge / Bernard Hirschel, Laurent Kaiser.	Hirschel, Bernard.		sida
100913031	999130313406676	SIDA : aspectos médico-legales y deontológicos / Marta Sanz, Javier Vega, Pelegrin Martínez.	Sanz Cid, Marta.		sida
101170729	9911707293406676	Sidà.			sida
101179213	9911792133406676	El SIDA : lineamientos sobre el manejo del síndrome de inmuno deficiencia-adquirida.			sida
101182898	9911828983406676	Sida : testimonio de una madre : dramático caso de la vida real / Rosa Feijoo Andrade			sida
101273166	9912731663406676	Le SIDA : réponses aux questions du personnel de santé.			sida
101437922	9914379223406676	SIDAI? : toda pregunta es importante / design [by] Katherine Shorey.			sida
8801878	996395873406676	Le SIDA / Michel Jossay, Yves Donadieu.	Jossay, Michel.		sida
8905055	996587133406676	SIDA : épidémies et sociétés, 20 et 21 juin 1987 : compte rendu de la réunion organisée aux Pensières à Veyrier-du-Lac, Annecy = AIDS : epidemics and societies / éditeur Charles Mérieux.			sida
8911614	996652363406676	Le S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	Cassuto, Jill-Patrice.		sida
9432215	997742393406676	SIDA.		SIDA (Paris, France)	sida
101457637	9914576373406676	Sidal.			sida

Title fields in Analytics

- Problem with Title (Normalized) including 245 \$b
- Title (Normalized) only returns two of the four titles shown below, but we need a tool that will return all four.

245 1 0 ‡a Structural bioinformatics : ‡b applications in preclinical drug discovery process / ‡c C. Gopi Mohan, editor.

245 0 0 ‡a Structural bioinformatics / ‡c edited by Jenny Gu, Philip E. Bourne.

245 0 0 ‡a Structural bioinformatics / ‡c edited by Philip E. Bourne, Helge Weissig.

245 1 0 ‡a Structural bioinformatics : ‡b methods and protocols / ‡c edited by Zoltán Gáspári.



Challenges with the Data

- Title (Normalized) and Title (Filing) don't meet our needs.
- Occasional extra spaces
- Initial Articles (in many languages)
- Accented Characters and Diacritics (in many styles)



Some Analytics functions we learned about

- TRIM(): Removes leading/trailing spaces
- REPLACE(): Replace one string with another
- LOWER(): Convert to lowercase
- RIGHT(): Return only the rightmost n characters of a string
- LENGTH(): Return the length of a string
- EVALUATE(): Run some non-Analytics Oracle SQL functions.



Using EVALUATE() for non-Analytics Functions

- TRANSLATE(): Replace specific character with another
- REGEXP_SUBSTR(): Returns a portion of a string, based on a regular expression
- REGEXP_REPLACE(): Returns a modified string, based on a regular expression



One Step at a Time

1. Remove initial articles
2. Convert to lower case
3. Isolate subfields \$a, \$n, and \$p
4. Normalize diacritics/accented Characters
5. Remove punctuation (and more diacritics)
6. Pack consecutive spaces
7. Trim leading/trailing spaces



Step 0: Start with “Title (Complete)”

- Title (Complete) is the best starting point.
 - Contains \$a, \$n and \$p, in order, with separators
- Title (Filing): subfields are in alpha order (a, b, c, n, p)
- Title (Normalized): subfields not separated with punctuation



Step 1: Remove initial articles

```
RIGHT("Bibliographic Details"."Title (Complete)",  
LENGTH("Bibliographic Details"."Title  
(Complete)") - (LENGTH(EVALUATE('REGEXP_REPLACE(%1,%2,%3,1,0,  
NULL) ',  
"Bibliographic Details"."Title (Complete)", '\s+', ' ')) -  
LENGTH(EVALUATE('REGEXP_REPLACE(%1,%2,%3,1,0,NULL) ',  
"Bibliographic Details"."Title (Filing)", '\s+', ' '))))
```

Title (Complete)	Title (Complete) step 1
Le S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.
Le SIDA / André Maghakian.	SIDA / André Maghakian.
Le SIDA / Michel Jossay, Yves Donadieu.	SIDA / Michel Jossay, Yves Donadieu.
Le SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.	SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.
Le sida.	sida.



Step 1a: Find the difference in length of Title (Complete) vs. Title (Filing)

```
LENGTH(EVALUATE('REGEXP_REPLACE(%1,%2,%3,1,0,NULL)',
"Bibliographic Details"."Title (Complete)",'\s+', ' ')) -
LENGTH(EVALUATE('REGEXP_REPLACE(%1,%2,%3,1,0,NULL)',
"Bibliographic Details"."Title (Filing)",'\s+', ' \'))
```

Title (Complete)	Title (Complete) step 1	Title (Complete) step 1a
Le S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	3
Le SIDA / André Maghakian.	SIDA / André Maghakian.	3
Le SIDA / Michel Jossay, Yves Donadieu.	SIDA / Michel Jossay, Yves Donadieu.	3
Le SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.	SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.	3
Le sida.	sida.	3
SIDA / A.K. Koffi.	SIDA / A.K. Koffi.	0



Step 1b: Subtract from length of Title (Complete)

LENGTH("Bibliographic Details"."Title (Complete)") - ([Expression from Step 1a])

Title (Complete)	Title (Complete) step 1	Title (Complete) step 1a	Title (Complete) step 1b
Le S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	3	70
Le SIDA / André Maghakian.	SIDA / André Maghakian.	3	24
Le SIDA / Michel Jossay, Yves Donadieu.	SIDA / Michel Jossay, Yves Donadieu.	3	36
Le SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.	SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.	3	76
Le sida.	sida.	3	5
SIDA / A.K. Koffi.	SIDA / A.K. Koffi.	0	18
SIDA / Cecilia Sepúlveda Carvajal, Alejandro Afani Saud, [editors].	SIDA / Cecilia Sepúlveda Carvajal, Alejandro Afani Saud, [editors].	0	68
SIDA / Institut de formation en soins infirmiers.	SIDA / Institut de formation en soins infirmiers.	0	49
SIDA / T. Sinare, M. Seroussi.	SIDA / T. Sinare, M. Seroussi.	0	30



Step 1c: Take characters from right side of Title (Complete)

RIGHT("Bibliographic Details"."Title (Complete)",
[Expression from Step 1b])

Title (Complete)	Title (Complete) step 1	Title (Complete) step 1a	Title (Complete) step 1b	Title (Complete) step 1c	Title (Filing)
Le S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	3	70	S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.
Le SIDA / André Maghakian.	SIDA / André Maghakian.	3	24	SIDA / André Maghakian.	SIDA / André Maghakian.
Le SIDA / Michel Jossay, Yves Donadieu.	SIDA / Michel Jossay, Yves Donadieu.	3	36	SIDA / Michel Jossay, Yves Donadieu.	SIDA / Michel Jossay, Yves Donadieu.
Le SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.	SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.	3	76	SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.	SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.
Le sida.	sida.	3	5	sida.	sida.
SIDA / A.K. Koffi.	SIDA / A.K. Koffi.	0	18	SIDA / A.K. Koffi.	SIDA / A.K. Koffi.
SIDA / Cecilia Sepúlveda Carvajal, Alejandro Afani Saud, [editors].	SIDA / Cecilia Sepúlveda Carvajal, Alejandro Afani Saud, [editors].	0	68	SIDA / Cecilia Sepúlveda Carvajal, Alejandro Afani Saud, [editors].	SIDA / Cecilia Sepúlveda Carvajal, Alejandro Afani Saud, [editors].
SIDA / Institut de formation en soins infirmiers.	SIDA / Institut de formation en soins infirmiers.	0	49	SIDA / Institut de formation en soins infirmiers.	SIDA / Institut de formation en soins infirmiers.
SIDA / T. Sinare, M. Seroussi.	SIDA / T. Sinare, M. Seroussi.	0	30	SIDA / T. Sinare, M. Seroussi.	SIDA / T. Sinare, M. Seroussi.
SIDA / [editado por] Cecilia Sepúlveda C., Alejandro Afani S.	SIDA / [editado por] Cecilia Sepúlveda C., Alejandro Afani S.	0	62	SIDA / [editado por] Cecilia Sepúlveda C., Alejandro Afani S.	SIDA / [editado por] Cecilia Sepúlveda C., Alejandro Afani S.



Step 1: Remove initial articles (completed)

```
RIGHT("Bibliographic Details"."Title (Complete)",  
LENGTH("Bibliographic Details"."Title  
(Complete)")) - (LENGTH(EVALUATE('regexp_replace(%1,%2,%3,1,0,  
NULL) '),  
"Bibliographic Details"."Title (Complete)", '\s+', '  
')) - LENGTH(EVALUATE('regexp_replace(%1,%2,%3,1,0,NULL) '),  
"Bibliographic Details"."Title (Filing)", '\s+', ' ')))))
```



Step 2: Convert to lower case

LOWER (*[Expression from Step 1]*)

Title (Complete)	Title (Complete) step 2
Le S.I.D.A. / Jill-Patrice Cassuto, Alain Pesce, Jean-François Quaranta.	s.i.d.a. / jill-patrice cassuto, alain pesce, jean-françois quaranta.
Le SIDA / André Maghakian.	sida / andré maghakian.
Le SIDA / Michel Jossay, Yves Donadieu.	sida / michel jossay, yves donadieu.
Le SIDA / sous la direction de Clément Olivier, Jean Robert et Réjean Thomas.	sida / sous la direction de clément olivier, jean robert et réjean thomas.
Le sida.	sida.
SIDA / A.K. Koffi.	sida / a.k. koffi.
SIDA / Cecilia Sepúlveda Carvajal, Alejandro Afani Saud, [editors].	sida / cecilia sepúlveda carvajal, alejandro afani saud, [editors].
SIDA / Institut de formation en soins infirmiers.	sida / institut de formation en soins infirmiers.
SIDA / T. Sinare, M. Seroussi.	sida / t. sinare, m. seroussi.
SIDA / [editado por] Cecilia Sepúlveda C., Alejandro Afani S.	sida / [editado por] cecilia sepúlveda c., alejandro afani s.



Step 3: Isolate subfields \$a, \$n, \$p

EVALUATE (REGEXP SUBSTR (%1, %2, 1, 1, NULL, 1) ', [Expression from Step 2]', '(.+?(\s:|\s/|\s=|. \$)) . *')

Title (Complete)	Title (Complete) step 3
Cancer cells in the circulating blood : a clinical study on the occurrence of cancer cells in the peripheral blood and in venous blood draining the tumour area at operation. [Tr. from the Danish].	cancer cells in the circulating blood :
Gan no jittai = [The facts of cancer / by Tomizo Yoshida and Toshio Kurokawa].	gan no jittai =
Gan to choju : gan wa seifuku dekiru = [Cancer and longevity; cancer can be conquered].	gan to choju :
IEEE transactions on circuits and systems. I, Regular papers : a publication of the IEEE Circuits and Systems Society.	ieee transactions on circuits and systems. i, regular papers :
Lung cancer: surgery and survival.	lung cancer: surgery and survival.
Nursing in cancer / Margaret Harrop, guest editor. The difficult patient: Vera M. Rubenstein, guest editor.	nursing in cancer /



Step 4b: Replace combining diacritics and others

```
REPLACE (REPLACE (REPLACE (REPLACE (REPLACE (
REPLACE (REPLACE (REPLACE (REPLACE (REPLACE (
REPLACE (REPLACE (REPLACE (REPLACE (REPLACE (
REPLACE ([Expression from Step 4a]
, 'æ', 'ae'), 'ǣ', 'ae'), 'Œ', 'oe'), 'œ', 'oe'), 'þ',
', 'th'), 'Ɔ', 'th'), 'īā', 'ia'), 't̄s̄', 'ts'), 'ī
ū', 'iu'), 'z̄h̄', 'zh'), '‘', '’'), '’', '’'), '’', '’'),
'̀', '̀'), '̀', '̀'), '˘', '˘')
```



Our progress so far (thru Step 4)...

Title (Complete)	Title (Complete) step 4
40 år med medicin.	40 ar med medicin.
50 cách đề phòng và chế ngự stress / Minh Anh, biên soạn.	50 cach de phong va che ngu stress /
50 års jubileum, 1923-1973. [Redaksjonskomite: Olav Hilmar Iversen og Per F. Marton.	50 ars jubileum, 1923-1973. [redaksjonskomite: olav hilmar iversen og per f. marton.
75 Jahre Bahnärztlicher Dienst, 1916-1991 / von J.P. Forney.	75 jahre bahnarztlicher dienst, 1916-1991 /
80 anni al servizio dell'umanità : Corpo delle infermiere volontarie, C.R.I., 1908-1988.	80 anni al servizio dell'umanita :
A Bajai városi tanács kórházának jubileumi emlékkönyve 1970. A kórház működésének 175. évében, 1795-1970 Szerk.: Cseh Imre.	bajai varosi tanacs korhazanak jubileumi emlekkönyve 1970. a korhaz mukodesenek 175. evében, 1795-1970 szerk.: cseh imre.
A Magyar Tudományos Akadémia másfél évszázada, 1825-1975 / írta, Kónya Sándor ... [et al.].	magyar tudomanyos akademia masfel evszazada, 1825-1975 /
A Magyar orvosi könyvtárügy és dokumentáció 20 éve. Szerkesztette: Szepesi Zoltánné.	magyar orvosi konyvtarugy es dokumentacio 20 eve. szerkesztette: szepesi zoltanne.
A Népegészségügyi Múzeum munkája.	nepegeszsegugyi muzeum munkaja.
A Terhesség szövődése belgyógyászati betegségekkel, írta Arvay Sándor [et al.].	terhesseg szovodese belgyogyaszati betegsegekkel, irta arvay sandor [et al.].
A acção biológica e terapêutica das radiações ...	accao biologica e terapeutica das radiacoes ...



Step 5: Remove punctuation (and more diacritics)

```
EVALUATE ('REGEXP_REPLACE' (%1, %2, %3, 1, 0, NULL) ',  
[Expression from Step 4], '^[:space:][:alnum:]', ''')
```

Title (Complete)	Title (Complete) step 5
40 år med medicin.	40 ar med medicin
50 cách đề phòng và chế ngự stress / Minh Anh, biên soạn.	50 cach de phong va che ngu stress
50 års jubileum, 1923-1973. [Redaksjonskomite: Olav Hilmar Iversen og Per F. Marton.	50 ars jubileum 19231973 redaksjonskomite olav hilmar iversen og per f marton
75 Jahre Bahnärztlicher Dienst, 1916-1991 / von J.P. Forney.	75 jahre bahnarztlicher dienst 19161991
80 anni al servizio dell'umanità : Corpo delle infermiere volontarie, C.R.I., 1908-1988.	80 anni al servizio dellumanita
A Bajai városi tanács kórházának jubileumi emlékkönyve 1970. A kórház működésének 175. évében, 1795-1970 Szerk.: Cseh Imre.	bajai varosi tanacs korhazanak jubileumi emlekkonyve 1970 a korhaz mukodesenek 175 eveben 17951970 szerk cseh imre
A Magyar Tudományos Akadémia másfél évszázada, 1825-1975 / írta, Kónya Sándor ... [et al.].	magyar tudomanyos akademia masfel evszazada 18251975
A Magyar orvosi könyvtárügy és dokumentáció 20 éve. Szerkesztette: Szepesi Zoltánné.	magyar orvosi konyvtarugy es dokumentacio 20 eve szerkesztette szepesi zoltanne
A Népegészségügyi Múzeum munkája.	nepegeszsegugyi muzeum munkaja
A Terhesség szövődése belgyógyászati betegségekkel, írta Arvay Sándor [et al.].	terhesseg szovodese belgyogyaszati betegsegekkel irta arvay sandor et al
A acção biológica e terapêutica das radiações ...	acciao biologica e terapeutica das radiacoes



Step 6: Pack consecutive spaces

```
EVALUATE (REGEXP_REPLACE (%1, %2, %3, 1, 0, NULL) ',  
[Expression from Step 5], '\s+', ' ')
```



Step 7: Trim leading/trailing spaces

```
TRIM(BOTH ' ' FROM [Expression from Step 8])
```



Institution's Folder

Location /Shared Folders/Community/Reports/Institutions

- Folders
 - My Folders
 - Shared Folders
 - Alma
 - Community
 - Reports
 - Consortia
 - Ex Libris Develop
 - Institutions**
 - 01ALLIANCE_
 - 01CARLI
 - 01SLCO_CUV
 - 01SUN-Stock
 - 01SUNV

Type	All	Sort	Name A-Z	Show More Details
	MSSA	Last Modified 11/19/2020 9:25:42 AM	Owner exlobi	Expand More
	Nasjonalbiblioteket	Last Modified 11/19/2020 9:25:42 AM	Owner exlobi	Expand More
	National Library of Medicine (U.S.)	Last Modified 8/30/2022 8:45:29 AM	Owner HE_958814950006676_6676_D_na91.alma.exlibrisgroup.com	Expand More
	NDUMD	Last Modified 11/19/2020 9:25:43 AM	Owner exlobi	Expand More
	New Mexico State University	Last Modified 11/19/2020 9:25:54 AM	Owner exlobi	Expand More
	Nord universitet	Last Modified 11/19/2020 9:25:57 AM	Owner exlobi	Expand More
	Norges idrettshøgskole	Last Modified 11/19/2020 9:25:58 AM	Owner exlobi	Expand More



Wrap up

- Existing functionality did not meet our needs
- Our formula works, but is slow (~5 minutes)
- IdeaExchange/NERS request
 - <https://bit.ly/3cyDiOS>



Resources

- Harvard University Library Technology Services Wiki
 - <https://bit.ly/3TvZVUW>
- ExLibris Developer Network: Tech Blog
 - <https://bit.ly/3AXZtr7>
- Oracle Database SQL Reference: SQL Functions
 - <https://bit.ly/3PYRpp>



Any questions???

Alvin Stockdale, MLIS

Librarian

National Library of Medicine

alvin.stockdale@nih.gov

Mike Davidson, MLS

Librarian

National Library of Medicine

mike.davidson@nih.gov



U.S. National Library of Medicine