



## 事例 NO.118

### PubMed2XL を利用した Keyword を含む PubMed データの抽出

#### ・質問

自分の研究領域の最近の動向を知るために、PubMed に掲載されている論文のデータを Keyword 項目込みで抽出したい。

[医学部保健学科学生，口頭，2018年7月]

#### ・調査の経緯

依頼の動機は，卒業研究を行う前に事前調査として，あるキーワードを持つ論文が毎年何本発表されているかを調べたいのだが，そのためのデータの抽出方法が判らないため，今回の相談に至ったとのことである。

#### 1. 既存の方法での検討

初めに検討したのは，当該研究領域をカバーするようなキーワードで検索した結果を PubMed の “Send to” メニューの “File” 形式でダウンロードする方法である。しかし，ダウンロードを行うと 1) データの改行位置が不規則 2) 項目ごとに区切り文字が挿入されていない 3) 項目が一部省略されている場合がある，という問題があった。このデータを調査に利用するためには，一件ずつリスト形式に修正する作業が必要となるため現実的ではないと思われた。

次に EndNote や Mendeley などの文献管理ツールの利

用を検討した。一旦 PubMed の検索結果のデータをこれらに取り込み，作成されたリストを MS-Word あるいは Excel への出力を試みた。しかし，データの改行の問題は解決されたものの，項目ごとの区切り文字は入らず，また Abstract や Keyword については出力できなかった。

#### 2. PubMed2XL の利用

さらに調査を進めると，Journal of the Medical Library Association 誌に “PubMed2XL (version 2.01)<sup>1)</sup>” という記事が掲載されていた。記事によると，1) PubMed から出力された XML ファイルを Excel ファイル (.xls) 等へ出力可能 2) Stylesheet を編集することで出力項目の変更が可能，とのことだったので，入手して試用してみることにした。ダウンロード及び操作方法については，上記記事で紹介されていた Web サイト，“blog.humaneguitarist.org<sup>2)</sup>” から入手した。

入手方法は “blog.humaneguitarist.org” (<http://blog.humaneguitarist.org/projects/pubmed2xl/>) にアクセスし，“Download” の項目にある “Click here to download the Windows self-installer (Windows)” の “Click here” をクリックするとインストーラーがダウンロードされる。このインストーラーを起動し “License Agreement” に同意するとインスールが開始され，パソコンのデスクトップ上に PubMed2XL のショートカットが作成される。このショー



# レファレンス事例集

トカットをクリックするとPubMed2XLが起動する。

操作手順は、1) 予めPubMedで検索した結果を“Send To”メニューからXML形式でダウンロードする 2) PubMed2XLを起動し、ToolsメニューからPubMed XML to Spreadsheetを選択する 3) ファイル選択画面が現れるので1)でダウンロードしたファイルを選択すると、“pubmed\_result\_(作成日時)”というExcelファイルが作成されるようになっている。

この結果、作成されたファイルが図1である。文献データがリスト化されており、項目毎にカンマで区切られているが、今回必要なKeywordの項目が含まれていなかった。そこで、Keywordの項目を出力できるようにStylesheetの編集を行うこととした。

Stylesheetの編集を行うには、“Tools”メニュー内の“Inspect Stylesheet”を選択する。ファイル選択画面が表示されるので“default”を右クリックし、“プログラムから開く”で提示されるプログラムからメモ帳やノートパッド等のエディターを選択する。初期設定のStylesheetが現れるので最後の行の“</spreadsheet>”の直前に図2の文を書き加え、別名で保存した。

編集したStylesheetをPubMed2XLの“Options”メ

ニュー内の“Change Stylesheet”の選択画面で選択し、再度実行した結果が図3である。リストの最後にKeywordが追加されるようになっている。

## ・回答

この時点で依頼者に連絡し、依頼者のパソコンにPubMed2XLと今回作成したStylesheetをインストールし、操作方法について簡単に説明したところ、依頼者からは大量のPubMedデータを簡単にリスト化できる点で期待通りの結果である、と喜ばれた。

## ・情報源

1) Isaak, D. PubMed2XL (version 2.01). [internet]. J Med Libr Assoc. 2016 Jan;104(1):92-4. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4722658/> [accessed 2018-09-27]

2) Arora N. PubMed2XL 2.01[internet]. <http://blog.humaneguitarist.org/projects/pubmed2xl/> [accessed 2018-09-27]

(弘前大学附属図書館医学情報グループ 藤井 真嗣 mfujii@hirosaki-u.ac.jp)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	
1	PMID	Digital Article	Abstract	Language	Journal	Journal	Journal	Journal	Journal	Journal	Journal	Journal	Journal	Citation	First Author	All Author	Affiliation	Related	Related	Related	Related	Journal	Article	Keyword	Abstract	Extractions	Extractions	Extractions
2	7002284	10.1186	A case of Primary s	eng	Surgica	Surg	C	101662	2018	Sep	4	1	121	Print	Kanako Kurata,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
3	7019944	10.1519	Very-Sho Yanaoka,	eng	Journal	J Strength		341508	2018	Sep	137	1	47.56	Internet	Takuma Yanaoka,	K	Graduat	click here	journal	Journal	Journal	Journal	Journal					
4	7002274	10.1186	Clinical This stu	eng	Docum	Doc Oph		037067	2018	07	16	1	103	Internet	Kentaro Kurata,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
5	7299134	10.1186	Optimis For pati	eng	BMC	in BMC	M	101190	2018	07	16	1	103	Internet	Tadashi Kato,	T	Aratam	click here	journal	Journal	Journal	Journal	Journal					
6	7299774	10.1007	invasive Scopuli	eng	Internat	Int J Her		311162	2018	Jul	58	7	4381.43	Internet	Keiji Kuri Kurata,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
7	7299795	10.2187	Efficacy The dose	eng	Artican	Artican		10298	2018	Jul	58	7	4381.43	Internet	Hitomi M Mori,	H	Departm	click here	journal	Journal	Journal	Journal	Journal					
8	7299791	10.2187	Cauma Epitheli	eng	Artican	Artican		10298	2018	Jul	58	7	4273.42	Internet	Makoto Kubo,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
9	7299412	10.1016	BRCAn The nei	eng	Clinical	Clin Bre		100887	2018	Jun	58	7	4273.42	Internet	Hitomi M Mori,	H	Departm	click here	journal	Journal	Journal	Journal	Journal					
10	7299791	10.1111	Body cdc The maj	eng	Internat	Int J Co		200716	2018	Aug	50	1	425.428	Internet	N Yoshio/Yoshioka	Kewple	C	click here	journal	Journal	Journal	Journal	Journal					
11	7298443	10.1038	Molecul Labor co	eng	Scientif	Sci Rep		101553	2018	May	8	1	8279	Internet	Katsushi Hosono,	I	Departm	click here	journal	Journal	Journal	Journal	Journal					
12	72978828		The Effe The purp	eng	Journal	J Sport		101174	2018	Jun	17	2	269.278	Internet	Takuma Yanaoka,	K	Graduat	click here	journal	Journal	Journal	Journal	Journal					
13	7297155	10.1016	Hieraci Recent	eng	Neurosci	Neurosci		700507	2018	Jul	382	1	127.143	Internet	Kiyoshi K Kurata,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
14	7297077	10.1588	The effe (Purpose)	eng	Journal	J Phys		105359	2018	Apr	30	1	625.629	Print	Yuka Hir Hamada,	K	Graduat	click here	journal	Journal	Journal	Journal	Journal					

図1. defaultのStylesheetによる出力結果

```

<column>
<title> keyword</title>
<!-- gets the abstract if it's in a single field like this: http://www.ncbi.nlm.nih.gov/pubmed/21476198 or whether it's
divided up into sections like this: http://www.ncbi.nlm.nih.gov/pubmed/21346227 -->
<hyperlink/>
<cell> <xsl:stylesheet version="1.0" encoding="UTF-8"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform"> <xsl:output method = "text" /> <xsl:template
match="/"> <xsl:for-each select="//Keyword"> <xsl:value-of select="." /> <xsl:text>,<
/></xsl:for-each> </xsl:template> </xsl:stylesheet> </cell>
</column>

```

図2. Keyword項目の追加のために作成したXML文  
この文を defaultのStylesheetに追加した

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	
1	PMID	Digital Article	Abstract	Language	Journal	Journal	Journal	Journal	Journal	Journal	Journal	Journal	Journal	Citation	First Author	All Author	Affiliation	Related	Related	Related	Related	Journal	Article	Keyword	Abstract	Extractions	Extractions	Extractions
2	7002284	10.1186	A case of Primary s	eng	Surgica	Surg	C	101662	2018	Sep	4	1	121	Print	Kanako Kurata,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
3	7019944	10.1519	Very-Sho Yanaoka,	eng	Journal	J Strength		341508	2018	Sep	137	1	47.56	Internet	Takuma Yanaoka,	K	Graduat	click here	journal	Journal	Journal	Journal	Journal					
4	7002274	10.1186	Clinical This stu	eng	Docum	Doc Oph		037067	2018	07	16	1	103	Internet	Kentaro Kurata,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
5	7299134	10.1186	Optimis For pati	eng	BMC	in BMC	M	101190	2018	07	16	1	103	Internet	Tadashi Kato,	T	Aratam	click here	journal	Journal	Journal	Journal	Journal					
6	7299774	10.1007	invasive Scopuli	eng	Internat	Int J Her		311162	2018	Jul	58	7	4381.43	Internet	Keiji Kuri Kurata,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
7	7299795	10.2187	Efficacy The dose	eng	Artican	Artican		10298	2018	Jul	58	7	4381.43	Internet	Hitomi M Mori,	H	Departm	click here	journal	Journal	Journal	Journal	Journal					
8	7299791	10.2187	Cauma Epitheli	eng	Artican	Artican		10298	2018	Jul	58	7	4273.42	Internet	Makoto Kubo,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
9	7299412	10.1016	BRCAnes The in	eng	Clinical	Clin Brea		100887	2018	Jun	58	7	4273.42	Internet	Hitomi M Mori,	H	Departm	click here	journal	Journal	Journal	Journal	Journal					
10	7299710	10.1111	Body cdc The maj	eng	Internat	Int J Co		200716	2018	Aug	50	1	425.428	Internet	N Yoshio/Yoshioka	Kewple	C	click here	journal	Journal	Journal	Journal	Journal					
11	7298443	10.1038	Molecul Labor co	eng	Scientif	Sci Rep		101553	2018	May	8	1	8279	Internet	Katsushi Hosono,	I	Departm	click here	journal	Journal	Journal	Journal	Journal					
12	72978828		The Effe The purp	eng	Journal	J Sport		101174	2018	Jun	17	2	269.278	Internet	Takuma Yanaoka,	K	Graduat	click here	journal	Journal	Journal	Journal	Journal					
13	7297155	10.1016	Hieraci Recent	eng	Neurosci	Neurosci		700507	2018	Jul	382	1	127.143	Internet	Kiyoshi K Kurata,	K	Departm	click here	journal	Journal	Journal	Journal	Journal					
14	7297077	10.1588	The effe (Purpose)	eng	Journal	J Phys		105359	2018	Apr	30	1	625.629	Print	Yuka Hir Hamada,	K	Graduat	click here	journal	Journal	Journal	Journal	Journal					

図3. Stylesheet編集後の出力結果  
項目の最後にKeywordが追加されている