



### 方法)茶筌(Ver.2.1)による形態素解析

WinCha  
ファイル(U) 編集(E) 解析(A) ヘルプ(H)

全文解析  
部分解析  
クリア

表層語  基本形  読み  発音  品詞  活用  区切り

オプション: -F "%m,%U(%P-%)%T,%F %n"

1-0-0-0..,未知語,,  
本章,名詞-一般,,  
の,助詞-連体化,,  
目的,名詞-一般,,  
は,助詞-係助詞,,  
医療,名詞-一般,,  
の,助詞-連体化,,  
成立,名詞-抽象接続,,  
における,助詞-格助詞-連語,,  
社会,名詞-一般,,  
資源,名詞-一般,,  
の,助詞-連体化,,  
必要,名詞-形容動詞語幹,,  
は,名詞-接尾-一般,,  
を,助詞-格助詞-一般,,

### 方法)Excelによる用語の抽出

合併処理表資料 - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M
	単語	名詞分類	記号分類	未知語分類	非自立語分類	品詞名	品詞b	品詞c	品詞d	品詞e	品詞f	品詞g	品詞h
1	1-0-0-0	1-0-0-0	0	0	0	1	1	1	1	0	0	0	未知語
2	2-1-0-0	2-1-0-0	1	0	0	1	1	1	1	1	1	1	一般
3	3-1-0-0	3-1-0-0	0	0	0	1	1	1	1	1	1	1	助詞
4	4-1-0-0	4-1-0-0	1	0	0	1	1	1	1	1	1	1	一般
5	5-1-0-0	5-1-0-0	0	0	0	1	1	1	1	1	1	1	助詞
6	6-1-0-0	6-1-0-0	0	0	0	1	1	1	1	1	1	1	記号
7	7-1-0-0	7-1-0-0	1	0	0	1	1	1	1	1	1	1	一般
8	8-1-0-0	8-1-0-0	0	0	0	1	1	1	1	1	1	1	助詞
9	9-1-0-0	9-1-0-0	1	0	0	1	1	1	1	1	1	1	連体化
10	10-1-0-0	10-1-0-0	1	0	0	1	1	1	1	1	1	1	一般
11	11-1-0-0	11-1-0-0	0	0	0	1	1	1	1	1	1	1	助詞
12	12-1-0-0	12-1-0-0	1	0	0	1	1	1	1	1	1	1	接助詞
13	13-1-0-0	13-1-0-0	1	0	0	1	1	1	1	1	1	1	一般
14	14-1-0-0	14-1-0-0	0	0	0	1	1	1	1	1	1	1	助詞
15	15-1-0-0	15-1-0-0	1	0	0	1	1	1	1	1	1	1	連体化
16	16-1-0-0	16-1-0-0	0	0	0	1	1	1	1	1	1	1	形容動詞
17	17-1-0-0	17-1-0-0	1	0	0	1	1	1	1	1	1	1	接尾
18	18-1-0-0	18-1-0-0	0	0	0	1	1	1	1	1	1	1	助詞
19	19-1-0-0	19-1-0-0	0	0	0	1	1	1	1	1	1	1	接助詞
20	20-1-0-0	20-1-0-0	0	0	0	1	1	1	1	1	1	1	記号

### 方法)Accessによるデータ集計

2004-2007年

用語	件数
日本医療研究推進財団	2
(社)全日本病院協会	2
(社)日本医療法人協会	2
(社)日本精神科病院協会	2
(社)日本病院会	3
100%負担	1
134条	1
16条	1
17条	1
1874年	2
1887年	1
1897年制定	1
10条	1
1906年	2
1915年	1
1922年	2
1925年	1
1927年	1
1933年	1
1938年	3
1942年	2
1946年	1
1946年制定	1

### 収載用語の変化の計測

●テキスト間の収載用語の変化の度合いは用語群の類似度によって測ることにした。

↓

Jaccard係数を評価指標に

## Jaccard係数 (Jaccard Similarity Coefficient)

- 集合の類似度(similarity)を測る指標  
⇒今日は2つの用語群の種類に注目した
  - 文章中の用語の出現頻度は考慮せず、用語の種類数によって類似度を測った。
- Jaccard係数( $J$ )の計算式
  - 用語群A・Bに含まれる用語の種類数を $a+b$ 、  
A・Bに共通に含まれる用語の種類数を $c$ とすると、
$$J = \frac{c}{a+b-c} \quad 0 \leq J \leq 1$$

参考:大垣俊一. 多様度と類似度、分類学的新指標. Argonauta 2008; 15: 10-22.

## 章別各年度版テキストの用語類似度 (数値:Jaccard係数)

第1章	2004年版	2007年版	2009年版	2012年版
2004年版		0.6793	0.5355	0.5000
2007年版			0.6783	0.6052
2009年版				0.6165
2012年版				

  

第2章	2004年版	2007年版	2009年版	2012年版
2004年版		0.7176	0.5092	0.4203
2007年版			0.6368	0.5093
2009年版				0.6440
2012年版				

## 章別各年度版テキストの用語類似度 (数値:Jaccard係数)

第1章	2004年版	2007年版	2009年版	2012年版
2004年版		0.6793	0.5355	0.5000
2007年版			0.6783	0.6052
2009年版				0.6165
2012年版				

  

第3章	2004年版	2007年版	2009年版	2012年版
2004年版		0.6872	0.1917	0.1924
2007年版			0.2274	0.2173
2009年版				0.7136
2012年版				

## 章別用語の平均類似度 (数値:Jaccard係数)

章	平均類似度	章題目
第1章	0.6025	医療管理総論
第2章	0.5729	医療管理各論 I
第3章	0.3716	医療管理各論 II
第4章	0.5165	医療情報学
第5章	0.4911	医療統計学
第6章	0.4792	診療情報管理論 I
第7章	0.4677	診療情報管理論 II

## 結果のまとめ

- 第1章から7章の類似度の平均  
⇒3章を除いて0.5前後を示した
- 我々が知るべき用語は、改訂によつてもそれほど大きくは変化していない可能性がある

## 考察(今後の課題)

- 専門用語の変化までは追えていない  
⇒専門用語を定義できなかつた  
(使える辞書がなかつた)
- 用語の正確な抽出には、辞書の充実が必要である。  
⇒今回得られたデータから独自辞書の作成が可能である