

2016 学際情報学概論 II

言葉のネットワーク

テキストデータマイニングの方法

担当：上田博人（文化人間情報学）

2016/04/05

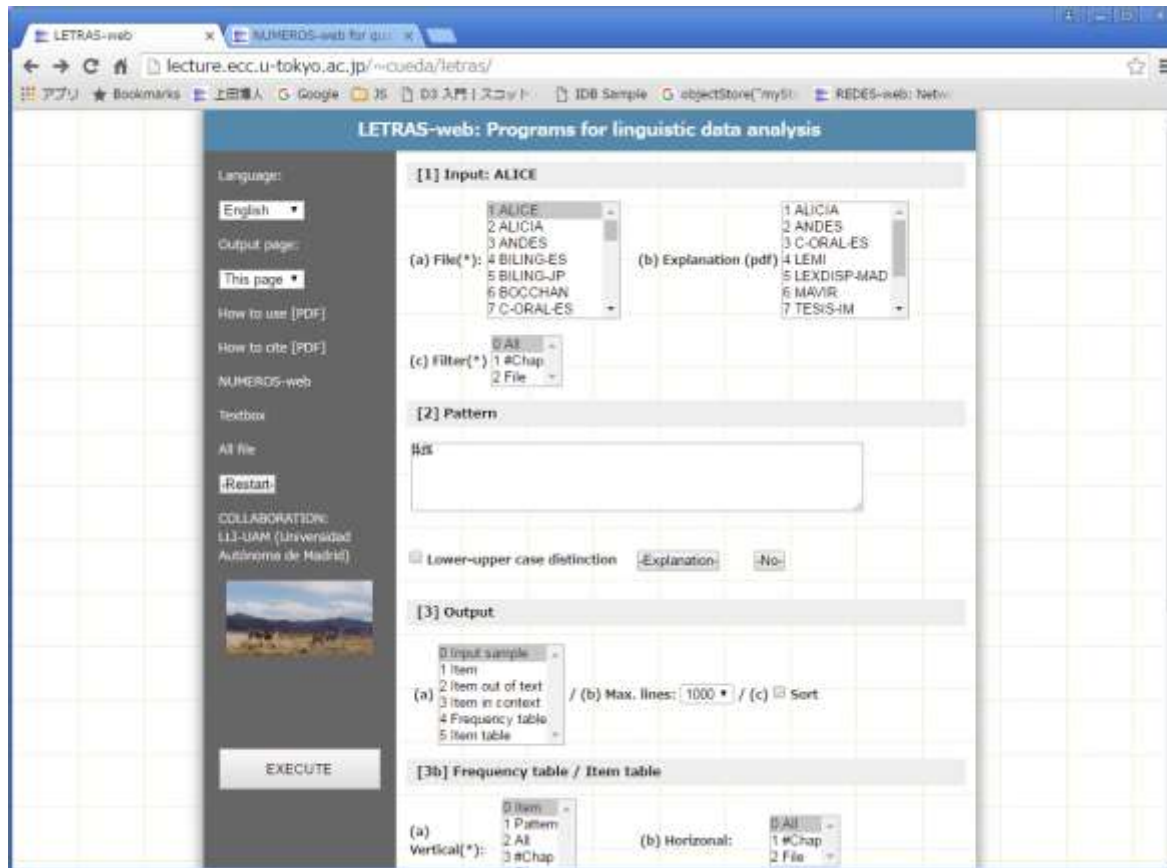
■ ネットワーク分析 =

● 言語データ分析：正規表現

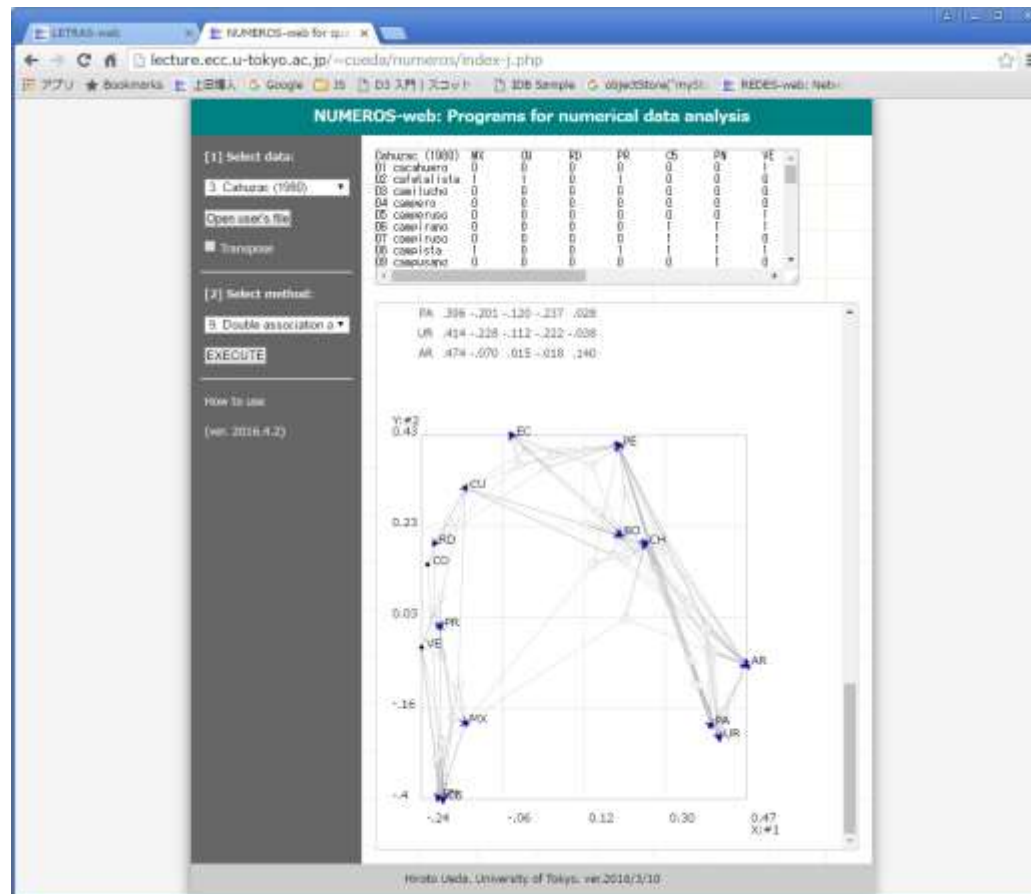
+

◆ 数量データ分析：連関分析

(+ 主成分分析)



<http://lecture.ecc.u-tokyo.ac.jp/~cueda/letras/>

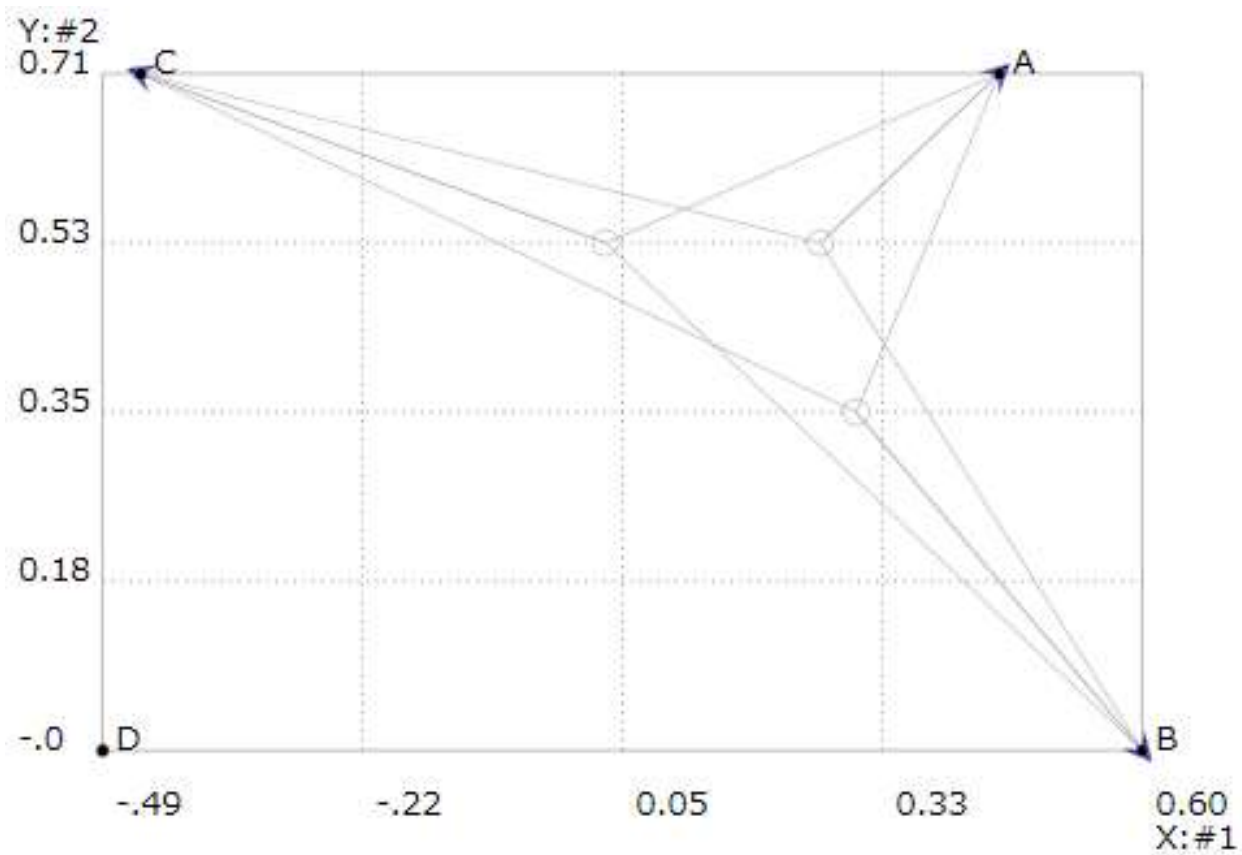


<http://lecture.ecc.u-tokyo.ac.jp/~cueda/numeros/index-j.php>

データ行列

D2	A	B	C	D
i1	1	1	0	0
i2	0	0	1	0
i3	0	1	0	0
i4	0	0	1	1
i5	1	1	1	0

ネットワーク



LETRAS-web NUMEROS-web for qual

lecture.ecc.u-tokyo.ac.jp/~cueda/numeros/index-j.php

NUMEROS-web: Programs for numerical data analysis

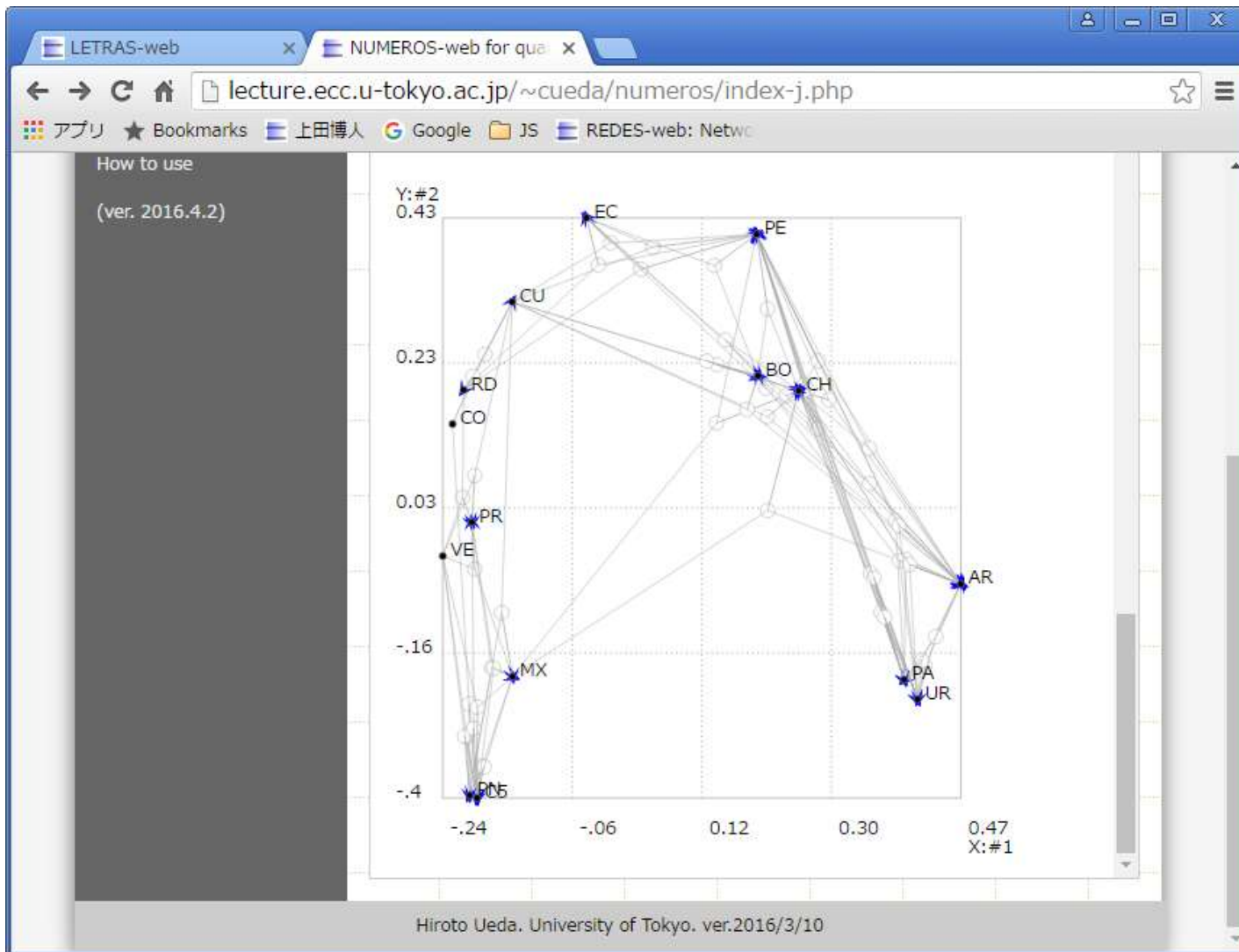
[1] Select data:
 3. Cahuzac (1980)
 Open user's file
 Transpose

Cahuzac (1980)	MX	CU	RD	PR	C5	PN	VE
01 cacahuero	0	0	0	0	0	0	1
02 cafetalista	1	1	0	1	0	0	0
03 camilicho	0	0	0	0	0	0	0
04 campero	0	0	0	0	0	0	0
05 camperuso	0	0	0	0	0	0	1
06 campirano	0	0	0	0	1	1	1
07 campiruso	0	0	0	0	1	1	0
08 campista	1	0	0	1	1	1	1
09 campusano	0	0	0	0	0	1	0

[2] Select method:
 9. Double association a
 EXECUTE

How to use
 (ver. 2016.4.2)

#	L. R.	LH > RH	L.s.	R.s.	L.p.	R.p.	Cooc.	Sup.	Cnf.	Lift	Synt.
1	13:15	14 PA:AR > UR	10	12	.213	.255	10	.213	1.000	3.917	.833
2	14:15	13 UR:AR > PA	12	10	.255	.213	10	.213	.833	3.917	.694
3	10:15	12 PE:AR > CH	6	5	.128	.106	5	.106	.833	7.833	.694
4	13:14	15 PA:UR > AR	10	16	.213	.340	10	.213	1.000	2.938	.625
5	12:15	10 CH:AR > PE	5	10	.106	.213	5	.106	1.000	4.700	.500
6	10:13	12 PE:PA > CH	2	5	.043	.106	2	.043	1.000	9.400	.400
7	10:14	12 PE:UR > CH	2	5	.043	.106	2	.043	1.000	9.400	.400
8	2:10	9 CU:PE > EC	3	8	.064	.170	3	.064	1.000	5.875	.375



尾崎隆 『ビジネスに活かすデータマイニング』

技術評論社(2014)

上田博人 『言語研究のためのテキスト分析法』

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengo/3-letras/letras-jp.pdf>

_____. 『言語研究のための数値データ分析法』

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengo/4-numeros/numeros.pdf>



検索: ueda hiroto → 「言語データ分析」 2. 3