

本期要目

壹、ROCLING 2011 專題演講簡介	第二~三頁
貳、專文-應用資料探勘於形聲字發音規則之探究(張嘉惠)	第四~十四頁
參、一〇一一年度會員資格更新暨個人資料異動通知單	第十五頁

ROCLING 2011

() ()

Google

ROCLING

ACL Anthology

(http://aclweb.org/anthology-new/)

http://sites.google.com/site/rocling2011/

7/1

1.

2.

3.

4.

1.

2.

)

7/1 7/31

http://www.aclclp.org.tw/doc/shipreg.htm

15

- ACM SIGIR 2011
- Collaborative Cyberporn Filtering with Collective Intelligence
- ()
- US\$600

ROCLING 2011 Keynotes



Title: Machine Transliteration - Translating the Untranslatables

Speaker: Dr. Haizhou Li, Institute of Infocomm, Singapore

Biography:

Dr. Haizhou Li is currently the Principal Scientist and Department Head of Human Language Technology at the Institute for Infocomm Research. Dr. Li has worked on speech and language technology in academia and industry since 1988. He taught in the University of Hong Kong (1988-1990), South China University of Technology (1990-1994), and Nanyang Technological University (2006-). He was a Visiting Professor at CRIN/INRIA in France (1994-1995), and at the University of New South Wales in Australia (2008). As a technologist, he was appointed as Research Manager in Apple-ISS Research Centre (1996-1998), Research Director in Lernout & Hauspie Asia Pacific (1999-2001), and Vice President in InfoTalk Corp. Ltd (2001-2003).

Dr. Li's research interests include automatic speech recognition, natural language processing and social robotics. He has published over 150 technical papers in international journals and conferences. He holds five international patents. Dr. Li now serves as an Associate Editor of IEEE Transactions on Audio, Speech and Language Processing, ACM Transactions on Speech and Language Processing, and Springer International Journal of Social Robotics. He is an elected Board Member of the International Speech Communication Association (ISCA, 2009-2013), an Executive Board Member of the Asian Federation of Natural Language Processing (AFNLP, 2006-2010), and a Senior Member of IEEE since 2001. Dr. Li was the Local Organizing Chair of SIGIR 2008 and ACL-IJCNLP 2009. He was appointed the General Chair of ACL 2012 and Interspeech 2014. He was the recipient of National Infocomm Award of Singapore in 2001. He was named one of the two Nokia Professors 2009 by Nokia Foundation in recognition of his contribution to speaker and language recognition technologies.

Abstract:

Machine transliteration is the process of automatically rewriting the script of a word from one language to another, while preserving pronunciation. The last decade has seen a tremendous progress and a growth of interests from theory to practice of machine transliteration. In this talk, I will present an overview of the fundamentals, algorithms and applications, in particular, transliteration between English and Chinese. I will also report the findings in the most recent transliteration evaluation campaigns - NEWS 2009 and NEWS 2010 Machine Transliteration Shared Tasks.



Title: Some Issues on Statistical Machine Translation Using Source and (or) Target Syntax

Speaker: Prof. Jingbo Zhu, Computer Science at the Northeastern University at Shenyang, China

Biography:

Prof. Jingbo Zhu is a full professor of Computer Science at the Northeastern University at Shenyang, China, and is in charge of research activities within the Natural Language Processing Laboratory (NEU-NLPlab, <http://www.nlplab.com>). He received his Ph.D. degree in computer software and theory from the Northeastern University in 1999. He was a visiting researcher at the City University of Hongkong (2004) and ISI, University of Southern California at Los Angeles (2006-2007), and was selected by the Program for New Century Excellent Talents in University, Ministry of Education (2005). His research interests include machine translation, syntactic parsing, sentiment analysis and text mining. He has published 100+ papers in many high-level journals and conferences including IEEE Transactions on Affective Computing, IEEE Transactions on Audio, Speech and Language Processing, ACM Transactions on Speech and Language Processing, ACM Transactions on Asian Language Information Processing, and ACL/EMNLP/Coling, etc.

Abstract:

Machine Translation (MT) is one of the oldest sub-fields in Natural Language Processing (NLP) and Artificial Intelligence (AI). During the last decade, syntax-based approaches have received growing interests in MT community, showing state-of-the-art performance for many language pairs such as Chinese-English. In this talk, I will present our recent work on syntax-based MT, and some approaches to performing translation using source and (or) target syntax, involving string-to-tree, tree-to-string and tree-to-tree SMT paradigms. Also, an empirical study is shown to compare the strengths and weaknesses among various syntax-based SMT approaches. Furthermore, several interesting issues are further addressed to investigate what the major problems in current (syntax-based) MT paradigm are. Finally, I will spend a little time to introduce a new open-source SMT toolkit (named NEUTrans) which was developed by the NLPLab of Northeastern University, and our current efforts on incorporating syntax-based SMT paradigms into this open SMT platform.



Title: To be determined

Speaker: Dr. Lee-Feng Chien, General Manager, Google, Taiwan

Abstract:

N/A yet

Biography:

N/A yet

2008

()

2009

11

1

2

3

4

5

“

”

Go!
Campus

2010
Saratoga High School

26

2010

AP

Easy
IWiLL
2010

1993

[10]

[2,3]

“

“

1
2
3
4
5

“

“

“

“

“

“

“

2009/11/10

2011/4/30

14309

9593

90.9%

“

“

“

“

Constrained Optimization

T

$w \in T$ w

$$\varepsilon_i \geq 0 \quad \sum_i \varepsilon_i^p$$

$$\varepsilon_i \quad 0$$

$$\varepsilon_i$$

0

A

$P_I(A) \quad P_V(A) \quad P_T(A)$

w w B

$P_I(B) \quad P_V(B) \quad P_T(B)$

$P_I(B) \quad P_I(A) \quad P_V(B) \quad P_V(A) \quad P_T(B) \quad P_T(A)$

$KL(P_I(B)||P_I(A)) \quad w$

$KL(P_V(B)||P_V(A)) \quad w \quad KL(P_T(B)||P_T(A)) \quad w$

1

90.90%

90.70%

244

20 40

90.03%

40

500 1000

1

1.		8721	858	14	90.90
2.		8701	648	244	90.70
		8576	575	442	89.39
3.		8910	683	0	92.80
		9362	231	0	97.50
		9207	386	0	95.90
	+ +	9413	180	0	98.10

92.8% 97.5% 95.9%

180 98.1% 9593 9413

2010

(support) (confidence)

(Stroke)

L16 16 s11 11 12-15

2

(60%~100%) Apriori 1% 0.5% weka 0.3% 3

“

“

No.				
1	INITIAL		{ \emptyset , , , , , }	
2	FINAL		{ \emptyset , , , , , }	
3	TONE		{1, 2, 3, 4, 5}	4
4	CONN		{ , , , , }	
5	RC			
6	PC			
7	PCLOC		{ , , , , , }	
8	PCI		{ \emptyset , , , , , }	
9	PCF		{ \emptyset , , , , , }	
10	PCT		{1, 2, 3, 4, 5}	1
11	WS		{L16, 14-15, s11}	s11
12	PCS		{L16, 14-15, s11}	s11
13	RCS			
14	WS-PCS		{s2, 4-5, L6}	s2
15	IU		{false , true}	false
16	FU		{false , true}	true
17	TU		{false , true}	false

3

Conf / Sup	60%	70%	80%	90%	100%
1%	304,330	217,346	143,301	87,324	50,054
0.5%	1,573,613	1,149,779	802,029	500,708	314,523
0.3%	6,625,518	5,144,742	3,879,619	2,809,951	1,810,585

4

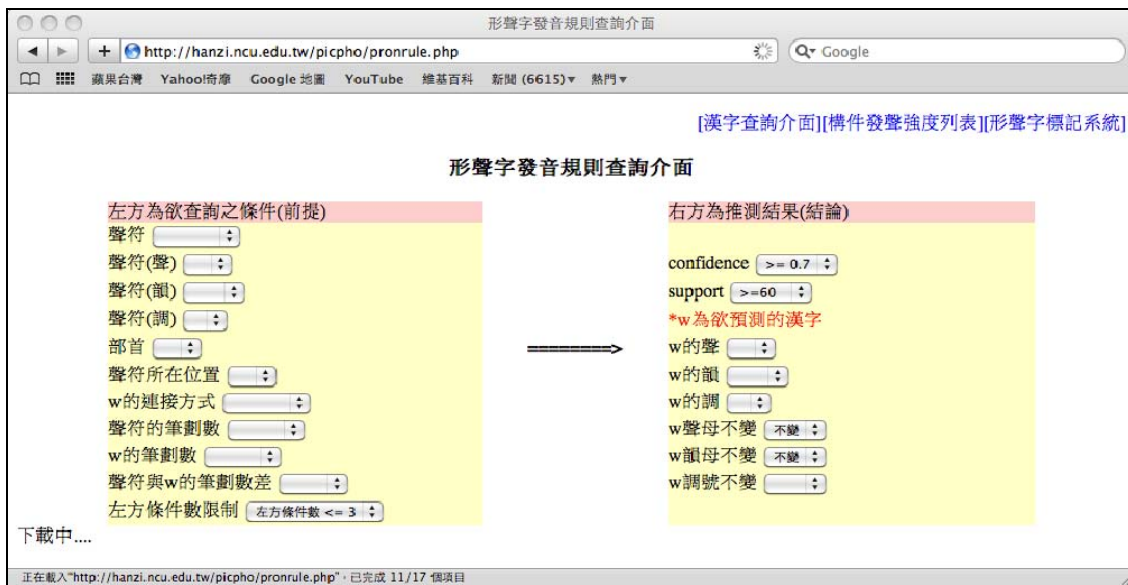
Conf / Sup	60%	70%	80%	90%	100%
1%	13,470	6,340	1,889	505	42
0.5%	61,171	32,089	15,243	7,561	5,190
0.3%	368,810	272,957	195,735	152,152	106,740

1

20

Ajax

...



1

(<http://hanzi.ncu.edu.tw/picpho/pronrule.php>)

w

>= 70%

>= 60

2

[]

[]

[=]

=]



2

3%

80%

15

3

R1-R3

16

(=)

3

11

- (R1) = (supp:197) =====> = (supp:178, conf:0.9) [], []
- (R2) = (supp:128) =====> = (supp:105, conf:0.82) [], []
- (R3) w =L16, =L16 (supp:123) =====> = (supp:98, conf: 0.8)

形聲字查詢系統

http://hanzi.ncu.edu.tw/picpho/lock_up_detail.php?SQL=PC_TH='8' and (RHS_TH<>'8' or RHS_T

條件:PC_TH='8' and (TH<>'8' or TH_changed<>'0') and `注音` is not null

字碼	是否為常用字	部首	注音	PC	聲符的聲母	聲符的韻母	聲符的調	w的筆劃數	聲符的筆劃數	w與聲符筆劃數差值	w的连接符號	聲符所在位置
泣	是	水	ㄑㄨㄣˋ	立	ㄌ	一	4	8	5	3	左右連接	右
翌	是	羽	ㄩˋ	立	ㄌ	一	4	11	5	6	上下連接	下
使	是	人	ㄕㄩˇ	吏	ㄌ	一	4	8	6	2	左右連接	右
莒	是	艸	ㄑㄩˇ	呂	ㄌ	ㄩ	3	11	7	4	上下連接	下
娘	是	女	ㄋㄩˊ	良	ㄌ	一	2	10	7	3	左右連接	右
焚	是	火	ㄈㄢˇ	林	ㄌ	一	2	12	8	4	上下連接	上
蔡	是	示	ㄘㄞˋ	林	ㄌ	一	2	13	8	5	上下連接	上
達	是	辵	ㄊㄨㄚˊ	奎	ㄌ	一	4	12	8	4	包圍式	內
剝	是	刀	ㄆㄠˊ	泉	ㄌ	ㄨㄢ	4	10	8	2	左右連接	左
數	是	攴	ㄕㄨˇ	婁	ㄌ	ㄨ	2	15	11	4	左右連接	左
膠	是	月	ㄑㄩㄟ	膠	ㄌ	一	4	15	11	4	左右連接	右
繆	是	系	ㄇㄩˋ	繆	ㄌ	一	4	17	11	6	左右連接	右
爍	是	火	ㄌㄞˋ	樂	ㄌ	ㄞ	4	19	15	4	左右連接	右
藥	是	艸	ㄩㄞˋ	樂	ㄌ	ㄞ	4	19	15	4	上下連接	下
鏢	是	金	ㄌㄞˋ	樂	ㄌ	ㄞ	4	23	15	8	左右連接	右
獺	是	犬	ㄊㄞˋ	賴	ㄌ	ㄞ	4	19	16	3	左右連接	右
龐	是	藍	ㄆㄨㄥˊ	龍	ㄌ	ㄨㄥ	2	19	16	3	上下連接	下
龐	是	龍	ㄆㄨㄥˊ	龍	ㄌ	ㄨㄥ	2	19	16	3	包圍式	內
灑	是	水	ㄕㄞˋ	麗	ㄌ	一	4	22	19	3	左右連接	右
旡		方	ㄈㄨㄟ	令	ㄌ	一	4	11	5	6	包圍式	內
翊		羽	ㄩˋ	立	ㄌ	一	4	11	5	6	左右連接	左
筵		竹	ㄑㄩㄟ	呂	ㄌ	ㄩ	3	13	7	6	上下連接	下
悝		心	ㄕㄞˋ	里	ㄌ	一	3	10	7	3	左右連接	右
裡		手	ㄌㄩˇ	里	ㄌ	一	3	10	7	3	左右連接	右
輪		目	ㄌㄨㄥˊ	侖	ㄌ	ㄨㄥ	2	13	8	5	左右連接	右

3

100% 0.5% 34

5 R4 R8

R7

11

16

R8

11

17

12

(R4) = , =3 , w =s3 (supp:16)
 =====> = , = (supp:16, conf:1) [], []

(R5) = , =2 , = , w =12-15 (supp:17)
 =====> = , = (supp:17, conf:1) [], []

(R6) = , =L16, w =4-5 (supp:16)
 =====> = , = (supp:16, conf:1) [], []

(R7) = , =1 , w =12-15 , =s11 (supp:16)
 =====> = , = (supp:16, conf:1) [], []

(R8) = , =1 , w = , w =s11 (supp:17)
 =====> = , = (supp:17, conf:1) [], []

•

• 7000 246 5631 1325
 246 54 4898
 87%(
 1993)

•

- [1] , , , 1988
- [2] , , , , ,
, 2005
- [3] , , , , ,
, 2008
- [4] . 2007
http://r9.ntue.edu.tw/activity/multiculture_conference/memoirs.html
- [5] 2009 2011
- [6] 2009-2010
- [7] 2010
- [8] , , , , ,
ROCLING XXI 2010
- [9] 403-448
- [10] 1993
- [11]
- [12] 2004
- [13] 2000
- [14]

E-mail _____

1. 10,000.- 1,000.- 500.-
2. 19166251

3. 100 7 31

101 100 7 1 101 6 30

_____ ()

VISA CARD MASTER CARD JCB CARD

_____(M/Y) _____

11529 2 128
(02)2788-3799 1502 (02)2788-1638
E-mail acclcp@hp.iis.sinica.edu.tw

