NTCIR Workshop 3
CLIR Task: General Statistics

Tracks		Total	SLIR				BLIR									MLIR								
# of Runs		189	110				50									29								
# of Groups		23	22				14									7								
Group ID	Nation	Submitted Runs	C-C	E-E	J-J	K-K	C-E	C-J	E-C	E-J	E-K	J-C	J-E	K-C	K-E	C-CE	C-CJ	C-JE	C-CJE	E-CE	E-CJE	E-JE	J-JE	J-CJE
APL	USA	23	2	3	3	3			3	3	3										3			
BRKLY	USA	11	1	3	1	1		1	1	1									1		1			
CECIR	Korea	6	1	1										2	2									
CMU	USA	7	3	1					3															
CRL	Japan	10	3	1	3	3																		
FJUIR	Taiwan	6	2		2	2																		
HKPU	Hong Kong	3	3																					
HUM	Canada	12	3	3	3	3																		
IFLAB	Japan	4		1	1								1										1	
ISCAS	China	34	3	3	2		3	3	3			2				3	3	3	3	3				
KRDLNLU	Singapore	1	1																					
KUNLP	Korea	4				1					3													
MSRA	China	6		3					3															
NTU	Taiwan	4		1																				3
OASIS	Japan	8		4	4																			
OKSAT	Japan	3		1	1							1												
PIRCS	USA	10	3	1					3											3				
POSTECH	Korea	9	3		3	3																		
SSTUT	Japan	6	1		1	1				1												1	1	
TJU	China	2										2												
TLRRD	USA	9	2	1	3					3														
TSB	Japan	8		2	3					3														
WATERLOO	Canada	3	3																					
# of Runs		189	34	29	30	17	3	4	16	11	6	5	1	2	2	3	3	3	4	6	4	1	2	3
# of Groups		23	15	15	13	8	1	2	6	5	2	3	1	1	1	1	1	1	2	2	2	1	2	1

NTCIR Workshop 3
CLIR Task: Run List

Tracks	Run Types	# of Runs	# of Groups	Total of # of Runs	Total of # of Groups
SLIR	C-C	34	15	110	22
	E-E	29	14		
	J-J	30	13		
	K-K	17	8		
BLIR	C-E	3	1	50	14
	C-J	4	2		
	E-C	16	6		
	E-J	11	5		
	E-K	6	2		
	J-C	5	3		
	J-E	1	1		
	K-C	2	1		
	K-E	2	1		
MLIR	C-CE	3	1	29	7
	C-CJ	3	1		
	C-JE	3	1		
	C-CJE	4	2		
	E-CE	6	2		
	E-CJE	4	2		
	E-JE	1	1		
	J-JE	2	2		
	J-CJE	3	1		
Total		189	23	189	23

NTCIR Workshop 3
CLIR Task: Techniques Description

Run ID	Index Unit	Index Tech	Index Struc	Query UnIt	Query Method	IR Model	Ranking	Query Expan	Trans Tech	Training Corpus
APL-C-C-TDNC-01	character n-grams of lengths 1 and 2	no morphological operations	Inverted file and corresponding dual file	n-grams generated from input text	automatic	statistical language model	TF*IDF	*1	No	No
APL-C-C-D-02	character n-grams of lengths 1 and 2	no morphological operations	Inverted file and corresponding dual file	n-grams generated from input text	automatic	statistical language model	TF*IDF	*1	No	No
APL-E-C-TDNC-01	character n-grams of lengths 1 and 2	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator	*5
APL-E-C-D-02	character n-grams of lengths 1 and 2	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator	*5
APL-E-C-T-03	character n-grams of lengths 1 and 2	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator	*5
APL-E-CEJ-TDNC-01	words (E); character n-grams of lengths 1, 2, and 3 (J), n-grams of lengths 1 and 2 (C)	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator.	*6
APL-E-CEJ-D-02	words (E); character n-grams of lengths 1, 2, and 3 (J), n-grams of lengths 1 and 2 (C)	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator.	*6
APL-E-CEJ-T-03	words (E); character n-grams of lengths 1, 2, and 3 (J), n-grams of lengths 1 and 2 (C)	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator.	*6
APL-E-E-TDNC-01	words	no morphological operations; whitespace-delimited, lower-cased tokens	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*1	No	

APL-E-E-D-02	words	no morphological operations; whitespace-delimited, lower-cased tokens	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*1	No	
APL-E-E-T-03	words	no morphological operations; whitespace-delimited, lower-cased tokens	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*1	No	
APL-E-J-TDNC-01	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator	*4
APL-E-J-D-02	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator	*4
APL-E-J-T-03	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator	*4
APL-E-K-TDNC-01	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator	*3
APL-E-K-D-02	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator	*3
APL-E-K-T-03	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	words	automatic	statistical language model	TF*IDF	*2	Single term translations were obtained by the Babelfish MT translator	*3
APL-J-J-TDNC-01	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	n-grams generated from input text	automatic	statistical language model	TF*IDF	*1	No	No
APL-J-J-D-02	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	n-grams generated from input text	automatic	statistical language model	TF*IDF	*1	No	No
APL-J-J-T-03	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	n-grams generated from input text	automatic	statistical language model	TF*IDF	*1	No	No
APL-K-K-TDNC-01	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	n-grams generated from input text	automatic	statistical language model	TF*IDF	*1	No	
APL-K-K-D-02	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	n-grams generated from input text	automatic	statistical language model	TF*IDF	*1	No	

APL-K-K-T-03	character n-grams of lengths 1, 2, and 3	no morphological operations	Inverted file and corresponding dual file	n-grams generated from input text	automatic	statistical language model	TF*IDF	*1		
Brkly-C-C-D-01	character + bi-character	stopwords removed	inverted file	character + bi-character	automatic	logistic regression	*7	50 terms selected from 20 top-ranked documents after the initial retrieval.		
Brkly-C-CJE-D-01		Direct merging from Brkly-C-C-D-01, Brkly-C-J-D-01, and Brkly-C-E-D-01.								
Brkly-C-J-D-01	character + bi-character	Hiragana discarded	inverted file	character + bi-character	automatic	logistic regression	*7	post-translation, 50 terms selected from 20 top-ranked documents after the initial retrieval.	Corpus-based for translating Chinese to English; MT-based for translating English to Japanese.	Hong Kong News, FBIS corpus.
Brkly-E-C-D-01	character + bi-character	stopwords removed	inverted file	character + bi-character	automatic	logistic regression	*7	50 terms selected from 20 top-ranked documents after the initial retrieval.	MT. Word not translated by MT were looked up in a bilingual dictionary created from parallel corpora.	Hong Kong News, FBIS corpus.
Brkly-E-CJE-D-01		Direct merging from Brkly-E-C-D-01, Brkly-E-J-D-01, and Brkly-E-E-D-04.								
Brkly-E-E-C-01	words	stopwords removed, content-words stemmed.	inverted file	words	automatic	logistic regression	*7	30 terms selected from 20 top-ranked documents after the initial retrieval.		
Brkly-E-E-TDN-02	words	stopwords removed, content-words stemmed.	inverted file	words	automatic	logistic regression	*7	30 terms selected from 20 top-ranked documents after the initial retrieval.		
Brkly-E-E-D-03	words	stopwords removed, content-words stemmed.	inverted file	words	automatic	logistic regression	*7	30 terms selected from 20 top-ranked documents after the initial retrieval.		
Brkly-E-J-D-01	character + bi-character	Hiragana discarded	inverted file	character + bi-character	automatic	logistic regression	*7	50 terms selected from 20 top-ranked documents after the initial retrieval.	MT.	None
Brkly-J-J-D-01	character + bi-character	Hiragana discarded	inverted file	character + bi-character	automatic	logistic regression	*7	50 terms selected from 20 top-ranked documents after the initial retrieval.		
Brkly-K-K-D-01	character + bi-character	stopwords removed	inverted file	character + bi-character	automatic	logistic regression	*7	50 terms selected from 20 top-ranked documents after the initial retrieval.		
*CECIR-C-C-TDNC-01	bigram	none	inverted index file	bigram	automatic	vector space	2-poisson	none	none	none
*CECIR-E-E-TDNC-01	word	stemming and reducing stop words	inverted index file	word	automatic	vector space	2-poisson	none	none	none

*CECIR-K-CC-TDNC-01	bigram	none	inverted index file	bigram	automatic	vector space	2-poisson	none	Dictionary based translation with manual disambiguation	none
*CECIR-K-CC-TDNC-02	bigram	none	inverted index file	bigram	automatic	vector space	2-poisson	none	Dictionary based translation with disambiguation by using semantic hierarchy	none
*CECIR-K-EE-TDNC-06	word	stemming and stop word elimination	inverted index file	words with stemming and stop word elimination	automatic	vector space	2-poisson	none	Dictionary based translation with manual disambiguation	none
*CECIR-K-EE-TDNC-07	word	stemming and stop word elimination	inverted index file	words with stemming and stop word elimination	automatic	vector space	2-poisson	none	Dictionary based translation with automatic disambiguation and weighting based on statistical information	none
CMU-C-C-TDC-01	phrase	segmentation	inverted file	word	automatic	VSM	tf+idf	PRF	none	
CMU-C-C-TDNC-02	phrase	segmentation	inverted file	word	automatic	VSM	tf+idf	PRF	none	
CMU-C-C-D-03	phrase	segmentation	inverted file	word	automatic	VSM	tf+idf	PRF	none	
CMU-E-C-TDC-01	phrase	segmentation	inverted file	word	automatic	VSM	tf+idf	Machine translation, dictionary	none	
CMU-E-C-TDC-02	phrase	segmentation	inverted file	word	automatic	VSM	tf+idf	Machine translation, dictionary	none	
CMU-E-C-TDNC-03	phrase	segmentation	inverted file	word	automatic	VSM	tf+idf	Machine translation, dictionary	none	
CMU-E-E-TDC-01	word	Stop words, stemming	inverted file	word	automatic	VSM	tf+idf	PRF	none	
CRL-C-C-TDNC-01	all the patterns of strings	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng., term position	automatic feedback	nothing	IREX, NTCIR1, NTCIR2
CRL-C-C-D-02	all the patterns of strings	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng., term position	automatic feedback	nothing	IREX, NTCIR1, NTCIR2
CRL-C-C-TC-03	all the patterns of strings	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng., term position	automatic feedback	nothing	IREX, NTCIR1, NTCIR2
CRL-E-E-TDNC-01	all the words	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng., term position	automatic feedback	nothing	IREX, NTCIR1, NTCIR2
CRL-J-J-TDNC-01	all the patterns of strings	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng., term position	automatic feedback	nothing	IREX, NTCIR1, NTCIR2
CRL-J-J-TDNC-02	all the patterns of strings	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng.	automatic feedback	nothing	IREX, NTCIR1, NTCIR2
CRL-J-J-D-03	all the patterns of strings	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng., term position	automatic feedback	nothing	IREX, NTCIR1, NTCIR2
CRL-K-K-TDNC-01	all the patterns of strings	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng., term position	automatic feedback	nothing	IREX, NTCIR1, NTCIR2

CRL-K-K-TDNC-02	all the patterns of strings	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng.	automatic feedback	nothing	IREX, NTCIR1, NTCIR2
CRL-K-K-D-03	all the patterns of strings	morphology	suffix array	word	automatic	okapi model (probabilistic model)	tf/idf, doc. Leng., term position	automatic feedback	nothing	IREX, NTCIR1, NTCIR2
FJUIR-C-C-C-01	*8	*9	inverted file.	*8	automatic	vector space model	tf/idf, word leng., doc. Leng.	None	None	none
FJUIR-C-C-D-02	*8	*9	inverted file.	*8	automatic	vector space model	t tf/idf, word leng., doc. Leng.	None	None	none
FJUIR-J-J-C-01	*8	*9	inverted file.	*8	automatic	vector space model	t tf/idf, word leng., doc. Leng.	None	None	none
FJUIR-J-J-D-02	*8	*9	inverted file.	*8	automatic	vector space model	t tf/idf, word leng., doc. Leng.	None	None	none
FJUIR-K-K-C-01	*8	*9	inverted file.	*8	automatic	vector space model	t tf/idf, word leng., doc. Leng.	None	None	none
FJUIR-K-K-D-02	*8	*9	inverted file.	*8	automatic	vector space model	t tf/idf, word leng., doc. Leng.	None	None	none
HKPU-C-CC-TDN-01	Short Word + Bigram	Hybrid	Extensible Inverted File	Short Word + Bigram	automatic	Probabilistic	2-Poisson	No		
HKPU-C-CC-C-02	Short Word + Bigram	Hybrid	Extensible Inverted File	Short Word + Bigram	automatic	Probabilistic	2-Poisson	No		
HKPU-C-CC-CT-03	Short Word + Bigram	Hybrid	Extensible Inverted File	Short Word + Bigram	automatic	Probabilistic	2-Poisson	No		
HUM-C-C-D-01	*10	case normalization of Latin, old Kanji normalized to new	inverted file	overlapping n-grams for CJK, words for Latin	automatic	vector space	tf/idf, tf dampened, idf squared, document length importance 50%	none	N/A	N/A
HUM-C-C-TC-02	*10	case normalization of Latin, old Kanji normalized to new	inverted file	overlapping n-grams for CJK, words for Latin	automatic	vector space	tf/idf, tf dampened, document length importance 0%	none	N/A	N/A
HUM-C-C-TDNC-03	*10	case normalization of Latin, old Kanji normalized to new	inverted file	overlapping n-grams for CJK, words for Latin	automatic	vector space	tf/idf, tf dampened, document length importance 100%	none	N/A	N/A
HUM-E-E-D-01	words	case normalization of Latin	inverted file	words	automatic	vector space	tf/idf, tf dampened, idf squared, document length importance 50%	none	N/A	N/A
HUM-E-E-TC-02	words	case normalization of Latin	inverted file	words	automatic	vector space	tf/idf, tf dampened, document length importance 0%	none	N/A	N/A
HUM-E-E-TDNC-03	words	case normalization of Latin	inverted file	words	automatic	vector space	tf/idf, tf dampened, document length importance 100%	none	N/A	N/A

HUM-J-J-D-01	*10	case normalization of Hiragana, Katakana and Latin, old Kanji normalized to new	inverted file	overlapping n-grams for CJK, words for Latin	automatic	vector space	tf/idf, tf dampened, idf squared, document length importance 50%	none	N/A	N/A
HUM-J-J-TC-02	*10	case normalization of Latin, old Kanji normalized to new	inverted file	overlapping n-grams for CJK, words for Latin	automatic	vector space	tf/idf, tf dampened, document length importance 0%	none	N/A	N/A
HUM-J-J-TDNC-03	*10	case normalization of Latin, old Kanji normalized to new	inverted file	overlapping n-grams for CJK, words for Latin	automatic	vector space	tf/idf, tf dampened, document length importance 100%	none	N/A	N/A
HUM-K-K-D-01	*10	Han mapped to Hangul, case normalization of Latin, old Kanji normalized to new	inverted file	overlapping n-grams for CJK, words for Latin	automatic	vector space	tf/idf, tf dampened, idf squared, document length importance 50%	none	N/A	N/A
HUM-K-K-TC-02	*10	Han mapped to Hangul, case normalization of Latin, old Kanji normalized to new	inverted file	overlapping n-grams for CJK, words for Latin	automatic	vector space	tf/idf, tf dampened, document length importance 0%	none	N/A	N/A
HUM-K-K-TDNC-03	*10	Han mapped to Hangul, case normalization of Latin, old Kanji normalized to new	inverted file	overlapping n-grams for CJK, words for Latin	automatic	vector space	tf/idf, tf dampened, document length importance 100%	none	N/A	N/A
IFLAB-E-E-D-01	word	morphology, stemming, POS	inverted file	word	automatic	probabilistic model	the Okapi method	No	No	No
IFLAB-J-E-D-01	word	morphology, stemming, POS	inverted file	word	automatic	probabilistic model	the Okapi method	No	dictionary/corpus-based, select the top translation	the target documents were used to produce language models
IFLAB-J-J-D-01	word	morphology, stemming, POS	inverted file	word	automatic	probabilistic model	the Okapi method	No	No	No
IFLAB-J-JE-D-01	word	morphology, stemming, POS	inverted file	word	automatic	probabilistic model	the Okapi method	No	dictionary/corpus-based, select the top translation	the target documents were used to produce language models
ISCAS-C-C-TC-01	2-char	2-char	Inverted File	2-char	automatic	VSM	Tf/idf	No		No
ISCAS-C-C-TDNC-02	2-char	2-char	Inverted File	2-char	automatic	VSM	Tf/idf	No		No
ISCAS-C-C-D-03	2-char	2-char	Inverted File	2-char	automatic	VSM	Tf/idf	No		No
ISCAS-C-CE-TC-01	C/2-char E/word	E/ stemming	Inverted File	C/2-char E/word	automatic	VSM	Tf/idf	Yes	Dictionary+MT	Pre-translation query expansion
ISCAS-C-CE-TDNC-02	C/2-char E/word	E/ stemming	Inverted File	C/2-char E/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-CE-D-03	C/2-char E/word	E/ stemming	Inverted File	C/2-char E/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-CJE-TC-01	C/2-char J/word E/word	JE/ stemming	Inverted File	C/2-char J/word E/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-CJE-TC-02	C/2-char J/word E/word	JE/ stemming	Inverted File	C/2-char J/word E/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No

ISCAS-C-CJE-D-03	C/2-char JE/word	JE/ stemming	Inverted File	C/2-char JE/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-CJ-TC-01	C/2-char J/word	J/ stemming	Inverted File	C/2-char J/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-CJ-TDNC-02	C/2-char J/word	J/ stemming	Inverted File	C/2-char J/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-CJ-D-03	C/2-char J/word	J/ stemming	Inverted File	C/2-char J/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-E-TC-01	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No	Dictionary+MT	
ISCAS-C-E-TDNC-02	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-E-D-03	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-JE-TC-01	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-JE-TDNC-02	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-JE-D-03	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-J-TC-01	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-J-TDNC-02	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-C-J-D-03	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-E-CE-TC-01	C/2-char E/word	E/ stemming	Inverted File	C/2-cha E/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-E-CE-TDNC-02	C/2-char E/word	E/ stemming	Inverted File	C/2-char E/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-E-CE-D-03	C/2-char E/word	E/ stemming	Inverted File	C/2-char E/word	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-E-C-TC-01	2-char	2-char	Inverted File	2-char	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-E-C-TDNC-02	2-char	2-char	Inverted File	2-char	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-E-C-D-03	2-char	2-char	Inverted File	2-char	automatic	VSM	Tf/idf	No	Dictionary+MT	No
ISCAS-E-E-TC-01	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No		No
ISCAS-E-E-TDNC-02	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No		No
ISCAS-E-E-D-03	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No		No
ISCAS-J-C-TDNC-01	2-char	2-char	Inverted File	2-char	automatic	VSM	Tf/idf	No	MT	No
ISCAS-J-C-D-02	2-char	2-char	Inverted File	2-char	automatic	VSM	Tf/idf	No	MT	No
ISCAS-J-J-TC-01	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No		No
ISCAS-J-J-D-02	Word	Stemming	Inverted File	Word	automatic	VSM	Tf/idf	No		No
KRDLNLU-C-C-D-01	Word and term	term extraction	inverted file	word and term	automatic	VSM	Tf/idf	No		No
KUNLP-E-K-D-01	word	POS (Nouns)	Inverted File	word	automatic	Okapi-like Probabilistic Model	tf/idf, document length	No	dictionary based-query translation which selects top 2 Korean words of each English word	the Financial Times, 1991-1994 at TREC CD4 is used to pre-query expansion :Korea Economic Daily (1994) is used to construct translation model.

KUNLP-E-K-D-02	word	POS (Nouns)	Inverted File	word	automatic	Okapi-like Probabilistic Model	tf/idf, document length	pre-query expansion using Okapi query expansion model	dictionary based-query translation which selects top 2 Korean words of each English word	the Financial Times, 1991-1994 at TREC CD4 is used to pre-query expansion :Korea Economic Daily (1994) is used to construct translation model.
^KUNLP-E-K-D-03	word	POS (Nouns)	Inverted File	word	interactive	Okapi-like Probabilistic Model	tf/idf, document length	pre-query expansion using Okapi query expansion model	dictionary based-query translation which selects top 2 Korean words of each English word.	The Financial Times, 1991-1994 at TREC CD4 is used to pre-query expansion. Korea Economic Daily (1994) is used to construct translation model.
KUNLP-K-K-D-01	word	POS (Nouns)	Inverted File	word	automatic	Okapi-like Probabilistic Model	tf/idf, document length	No	n/a	n/a
MSRA-E-C-TDNC-01	character, word	stopword list	inverted file	character, word	automatic	probabilistic +model	bm2500	pseudo relevance feedback	dictionary based, decaying co-occurrence model, syntactic dependence relations	CIRB010, CIRB020, Wall Street Journal(1987-1992), People's Daily(1980-1998)
MSRA-E-C-TDNC-02	character, word	stopword list	inverted file	character, word	automatic	probabilistic model	bm2500	pseudo relevance feedback	dictionary based, decaying co-occurrence model, syntactic dependence relations	CIRB010, CIRB020, Wall Street Journal(1987-1992), People's Daily(1980-1998)
MSRA-E-C-D-03	character, word	stopword list	inverted file	character, word	automatic	probabilistic model	bm2500	No	dictionary based, decaying co-occurrence model, syntactic dependence relations	CIRB010, CIRB020, Wall Street Journal(1987-1992), People's Daily(1980-1998)
MSRA-E-E-TDNC-01	Word	using stoplist and stemming	inverted file	word	automatic	probabilistic model	bm2500	no query expansion	None	None
MSRA-E-E-TDNC-02	Word	using stoplist and stemming	inverted file	word	automatic	probabilistic model	bm2500	no query expansion	Although we didn't do query expansion, but we used the information we got from the first retrieval result	None
MSRA-E-E-TDNC-03	Word	using stoplist and stemming	inverted file	word	automatic	probabilistic model	bm2500	no query expansion	Using query expansion	None
*NTU-E-E-D-01	word	N/A	inverted file	word	automatic	Vector space model	tf*idf	None	None	TREC6 text collection
*NTU-J-CJE-D-01	C: haracter bigram, J, E: word	J: morphology	inverted file	C: haracter bigram, J, E: word	automatic	Vector space model	tf*idf	None	*11	ASBC corpus, TREC6 text collection
*NTU-J-CJE-D-02	C: haracter bigram, J, E: word	J: morphology	inverted file	C: haracter bigram, J, E: word	automatic	Vector space model	tf*idf	None	*11	ASBC corpus, TREC6 text collection

*NTU-J-CJE-D-03	C: haracter bigram, J, E: word	J: morphology	inverted file	C: haracter bigram, J, E: word	automatic	Vector space model	tf*idf	None	*11	ASBC corpus, TREC6 text collection
!OASIS-E-E-D-01	word	stop word were discarded.	Inverted index	word	automatic	vector space model	tf/idf	No	None	ntc-e02-mai98.txt
!OASIS-E-E-D-02	word	stop word were discarded	Inverted index	word	automatic	vector space model	tf/idf	*12	None	ntc-e02-mai98.txt
!OASIS-E-E-D-03	word	stop word were discarded	Inverted index	word	automatic	vector space model	tf/idf	No	None	ntc-e02-mai99.txt
#!OASIS-E-E-D-04	word	stop word were discarded	Inverted index	word	automatic	vector space model	tf/idf	*12	None	ntc-e02-mai99.txt
!OASIS-J-J-D-01	combination bi-word and phrases	*13	Inverted index	combination bi-word and phrases	automatic	vector space model	tf/idf	*12	None	ntc-j-mai-99.txt
!OASIS-J-J-D-02	combination bi-word and phrases	*13	Inverted index	combination bi-word and phrases	automatic	vector space model	tf/idf	*12	None	ntc-j-mai-98.txt
!OASIS-J-J-D-03	combination bi-word and phrases	*13	Inverted index	combination bi-word and phrases	automatic	vector space model	tf/idf	No	None	ntc-j-mai-98.txt
#!OASIS-J-J-D-04	combination bi-word and phrases	*13	Inverted index	combination bi-word and phrases	automatic	vector space model	tf/idf	No	None	ntc-j-mai-99.txt
^OKSAT-E-E-D-01	n-gram	n-gram	inverted -gram index in Tree.	word + phrase	interactive	probabilistic model	tf/idf, essential keywords	synonym	No	No
^OKSAT-J-C-D-01	n-gram	n-gram	inverted -gram index	word + phrase	interactive	probabilistic model	tf/idf	Expand query	Dictionary-based	Post-translation
OKSAT-J-J-D-01	n-gram	n-gram	inverted -gram index in Tree.	word + phrase	automatic	probabilistic model	tf/idf, essential keywords	synonym	No	No
pircs-C-C-D-001	2-gram + 1-gram	5 stopwrd ch;	invrt-file network	2-gram + 1-gram	automatic	probabilistic, activation-spreading	activation: tf, ictf, doc + qry length	Based on term freq, doc length	none	N/A
pircs-C-C-D-002	2-gram + 1-gram; short wd	5 stopwrd ch; dictionary	invrt-file network	2-gram + 1-gram; short wd	automatic	probabilistic, activation-spreading	activation: tf, ictf, doc + qry length + combine 2 retrievals	Based on term freq, doc length	none	N/A
pircs-C-C-TDNC-003	2-gram + 1-gram; short wd	5 stopwrd ch; dictionary	invrt-file network	2-gram + 1-gram; short wd	automatic	probabilistic, activation-spreading	activation: tf, ictf, doc + qry length + combine 2 retrievals	Based on term freq, doc length	none	N/A
pircs-E-C-D-001	short wrd + char	dictionary	invrt-file network	short wd + char	automatic	probabilistic, activation-spreading	activation: tf, ictf, doc + qry length	post trans expansion	biling. dict + MT software	N/A
pircs-E-C-D-002	short wrd + char	dictionary	invrt-file network	short wd + char	automatic	probabilistic, activation-spreading	activation: tf, ictf, doc + qry length	pre & post trans expan	biling. dict + MT software	N/A
pircs-E-C-TDNC-003	short wrd + char	dictionary	invrt-file network	short wd + char	automatic	probabilistic, activation-spreading	activation: tf, ictf, doc + qry length	post trans expansion	biling. dict + MT software	N/A
pircs-E-EC-D-001	stems + 2-wd phr; short wd	Porter stemming, stopwords; dictionary	invrt-file network	stems; short wd	automatic	probabilistic, activation-spreading	activation: tf, ictf, doc + qry length + English RSV directly vs. E-C-D-002	pre & post trans expan	biling. dict + MT software	N/A

pircs-E-EC-D-002	stems + 2-wd phr; short wd	Porter stemming, stopwords; dictionary	invrt-file network	stems; short wd	automatic	probabilistic, activation-spreading	English RSV adjusted vs. E-C-D-002	pre & post trans expan	biling. dict + MT software	N/A
pircs-E-EC-D-003	stems + 2-wd phr; short wd	Porter stemming, stopwords; dictionary	invrt-file network	stems; short wd	automatic	probabilistic, activation-spreading	English RSV adjusted vs. E-C-D-001	post trans expansion	biling. dict + MT software	N/A
pircs-E-E-TDNC-001	stems + 2-wd phr	Porter stemming, stopwords	invrt-file network	stems + 2-wd phr	automatic	probabilistic, activation-spreading	activation: tf, ictf, doc + qry length	Based on term freq, doc length	none	N/A
POSTECH-C-C-D-01	morpheme	POS tagging	inverted file	morpheme	automatic	probabilistic model	modified Okapi BM25	no query expansion	N/A	N/A
POSTECH-C-C-C-02	morpheme	POS tagging	inverted file	morpheme	automatic	probabilistic model	modified Okapi BM25	no query expansion	N/A	N/A
POSTECH-C-C-T-03	morpheme	POS tagging	inverted file	morpheme	automatic	probabilistic model	modified Okapi BM25	no query expansion	N/A	N/A
POSTECH-J-J-C-01	morpheme	POS tagging	inverted file	morpheme	automatic	probabilistic model	modified Okapi BM25	no query expansion	N/A	N/A
POSTECH-J-J-D-02	morpheme	POS tagging	inverted file	morpheme	automatic	probabilistic model	modified Okapi BM25	no query expansion	N/A	N/A
POSTECH-J-J-T-03	morpheme	POS tagging	inverted file	morpheme	automatic	probabilistic model	modified Okapi BM25	no query expansion	N/A	N/A
POSTECH-K-K-D-01	morpheme	POS tagging	inverted file	morpheme	automatic	probabilistic model	modified Okapi BM25	no query expansion	N/A	N/A
POSTECH-K-K-D-02	morpheme	POS tagging	inverted file	morpheme	automatic	probabilistic model	modified Okapi BM25	no query expansion	N/A	N/A
POSTECH-K-K-C-03	morpheme	POS tagging	inverted file	morpheme	automatic	probabilistic model	modified Okapi BM25	no query expansion	N/A	N/A
SSTUT-C-C-D-01	character	all	suffix array	bigram	automatic	probabilistic model	empirical term weighting	no query expansion	automatic extraction of translation word from parallel corpus	NTCIR1/NTCIR2 Corpus
SSTUT-E-J-D-01	character	all	suffix array	bigram	automatic	probabilistic model	empirical term weighting	no query expansion	automatic extraction of translation word from parallel corpus	NTCIR1/NTCIR2 Corpus
SSTUT-E-JE-D-01	character	all	suffix array	bigram	automatic	probabilistic model	empirical term weighting	no query expansion	automatic extraction of translation word from parallel corpus	NTCIR1/NTCIR2 Corpus
SSTUT-J-J-D-01	character	all	suffix array	bigram	automatic	probabilistic model	empirical term weighting	no query expansion	automatic extraction of translation word from parallel corpus	NTCIR1/NTCIR2 Corpus
SSTUT-J-JE-D-01	character	all	suffix array	bigram	automatic	probabilistic model	empirical term weighting	no query expansion	automatic extraction of translation word from parallel corpus	NTCIR1/NTCIR2 Corpus
SSTUT-K-K-D-01	character	all	suffix array	bigram	automatic	probabilistic model	empirical term weighting	no query expansion	automatic extraction of translation word from parallel corpus	NTCIR1/NTCIR2 Corpus
TJU-J-C-D-01	bi-charater	stemming	inverted file	bi-charater	automatic	vector space model	tf*idf	no query expansion	dictionary-based	N/A
TJU-J-C-TDNC-02	bi-charater	stemming	inverted file	bi-charater	automatic	vector space model	tf*idf	no query expansion	dictionary-based	N/A
tlrrd-C-C-DC-01	bi-character	none	inverted file	bi-character	automatic	inference networks, probabilistic model	N/A	none	N/A	N/A
tlrrd-C-C-D-02	bi-character	none	inverted file	bi-character	automatic	inference networks, probabilistic model	N/A	none	N/A	N/A
tlrrd-E-E-DC-01	word	stemming	inverted file	word	automatic	inference networks, probabilistic model	N/A	none	N/A	N/A

tlrrd-E-J-D-01	bi-character	none	inverted file	word for English, bi-character for Japanese	automatic	inference networks, probabilistic model	N/A	none	dictionary-based, select all	N/A
tlrrd-E-J-D-02	word	morphology, stemming	inverted file	word	automatic	inference networks, probabilistic model	N/A	none	dictionary-based, select all	N/A
tlrrd-E-J-DC-03	bi-character	none	inverted file	word for English, bi-character for Japanese	automatic	inference networks, probabilistic model	N/A	none	dictionary-based, select all	N/A
tlrrd-J-J-D-01	bi-character	none	inverted file	bi-character	automatic	inference networks, probabilistic model	N/A	none	N/A	N/A
tlrrd-J-J-D-02	bi-character	morphology, stemming	inverted file	bi-character	automatic	inference networks, probabilistic model	N/A	none	N/A	N/A
tlrrd-J-J-DC-03	bi-character	none	inverted file	bi-character	automatic	inference networks, probabilistic model	N/A	none	N/A	N/A
TSB-E-E-D-01	word	stemming	inverted file	word	automatic	probabilistic	BM25	expansion based on the offer weight	N/A	target docs only
TSB-E-E-D-02	word	stemming	inverted file	word	automatic	probabilistic	BM25	expansion based on chi-square	N/A	target docs only
TSB-E-J-D-01	morphemes	morphological analysis	inverted files	morphemes	automatic	probabilistic	tf-idf / tf-relevance weight	expansion based on document scores (combination of 2 runs)	MT	mainichi1998-1999 only
TSB-E-J-D-02	morphemes	morphological analysis	inverted files	morphemes	automatic	probabilistic	tf-idf / tf-relevance weight	expansion based on document scores/chi-square (combination of 3 runs)	MT	mainichi1998-2000
TSB-E-J-D-03	morphemes	morphological analysis	inverted files	morphemes	automatic	probabilistic	tf-idf / tf-relevance weight	expansion based on document scores/chi-square (combination of 3 runs)	MT	mainichi1998-2000
TSB-J-J-D-01	morphemes	morphological analysis	inverted files	morphemes	automatic	probabilistic	tf-idf / tf-relevance weight	expansion based on document scores (combination of 2 runs)	N/A	mainichi1998-1999
TSB-J-J-D-02	morphemes	morphological analysis	inverted files	morphemes	automatic	probabilistic	tf-idf / tf-relevance weight	expansion based on document scores/chi-square (combination of 3 runs)	N/A	mainichi1998-2000
TSB-J-J-D-03	morphemes	morphological analysis	inverted files	morphemes	automatic	probabilistic	tf-idf / tf-relevance weight	expansion based on document scores/chi-square (combination of 3 runs)	N/A	mainichi1998-2000
WATERLOO-C-C-TDNC-01	Word	Self-supervised word segmentation	Inverted files	word	automatic	Vector space model	tf, qtf, doc. leng.	N/A	None	None
WATERLOO-C-C-C-02	character	none	Inverted files	word	automatic	Vector space model	tf, qtf, doc. leng.	N/A	None	None

WATERLOO-C-C-D-03	word	Self-supervised word segmentation	Inverted files	word	automatic	Vector space model	tf, qtf, doc. leng.	N/A	None	None

CECIR- and *NTU-* are runs submitted from the members of Executive Committee of CLIR Task, NTCIR Workshop 3.
#OASIS-E-E-D-04 and #OASIS-E-E-D-04 are additional runs (just for reference).
^KUNLP-E-K-D-03, ^OKSAT-E-E-D-01, and ^OKSAT-J-C-D-01 are interactive runs.
!OASIS-E-E-D-01, !OASIS-E-E-D-02, !OASIS-E-E-D-03, !OASIS-E-E-D-04, !OASIS-J-J-D-01, !OASIS-J-J-D-02, !OASIS-J-J-D-03, and !OASIS-J-J-D-04 uses part of document set. Their results could not be used to compare to other runs directly. The details could be referred to the corresponding paper in workshop proceedings

*1: Query was expanded to 60 weighted terms. Additional terms were weighted and selected based on their frequencies in the query, in the top 20 documents, in the bottom 75 (of 1000) documents, and the overall frequency in the collection.
*2: Query was expanded to 60 weighted words prior to translation. The additional words were weighted and selected based on their frequencies in the query, in the top 20 ranked English documents, in the bottom 75 (of 1000) English documents, and the overall frequency in the English collection. No post-translation expansion was applied.
*3: The English subcollection of the NTCIR-3 CLIR document collection was used to assist with pre-translation expansion. When an English term could not be translated, it waseft in the query. Non word-spanning character n-grams were produced from the resultant query after translation; these n-grams were the query used to search against the Korean collection.
*4: The English subcollection of the NTCIR-3 CLIR document collection was used to assist with pre-translation expansion. When an English term could not be translated, it waseft in the query. Non word-spanning character n-grams were produced from the resultant query after translation; these n-grams were the query used to search against the Japan collection.
*5: The English subcollection of the NTCIR-3 CLIR document collection was used to assist with pre-translation expansion. When an English term could not be translated, it waseft in the query. Non word-spanning character n-grams were produced from the resultant query after translation; these n-grams were the query used to search against the Chinese collection.
*6: The English subcollection of the NTCIR-3 CLIR document collection was used to assist with pre-translation expansion for the Chinese and Japanese queries. When an English term could not be translated, it was left in the query. Non word-spanning character n-grams were produced from the resultant query after translation; these n-grams were the query used to search against the Chinese/Japanese collection. Pre-translation expansion alone was used when searching the Chinese and Japanese collections without additional blind relevance feedback, but relevancefeedback was used for the search against the English collection.
*7: qtf (within-query term fequency), dtf (within-document term frequency), ctf (within-collection term frequency), ql (query length), dl (document length), cl (collection length)
*8: 1-gram, 2-gram, dictionary-based indexing, and key-phrase formulation. Thus the index terms include single characters, all 2-grams, dictionary words, and repeated words/phrases unregistered in the dictionary.
*9: a phrase formulation technique based on repeated sequence extraction to combine unsegmented characters or words into new words or phrases.
*10: n-grams for CJK (after conversion to canonical Unicode (UTF-16)), words for Latin
*11: We used CO Model to translate queries. This model uses word co-occurrence, information to select best translation. We select Japanese topics as original queries. The original Japanese topics were translated into English. Then the translated English topics were translated into Chinese.
*12: Every query was processed twice: First search generated 2 docs. Words consisted of 3 characters were considered as candidates for expansion. Words, which occurred four times, were selected from this set. Their number did not exceed a half of the word number in the original query. In the case of the necessity, the random selection were utilized
*13: Katakana sequences were considered as a word; hiragana characters were discarded; from the rest of the text overlapping bi-gram were taken into account