

NTCIR Workshop Ad Hoc IR Task

	Group's ID	
	Run ID	
	1.Indexing	
	1.1.Indexing	
1.1.1	1)Index units for Japanese	1.uni-gram 2.bi-gram 3.other n-gram 4.word 5.phrase 6.other 0.NA (answerd) 99.NA
1.1.2	2)Index units for English	1.n-gram 2.word 3.phrase 4.other 0.NA (answerd) 99.NA
1.1.3	3)Index units for E-term in J	1.n-gram 2.word 3.phrase 4.other 0.NA (answerd) 99.NA
1.1.4	4)Method Indexing	1.lexicon 2.morphological analysis 3.other 0.NA (answerd) 99.NA
1.1.5	5)Method Selection index term	1.stop word 2.type of characters 3.part of speech 4.other 0.NA (answerd) 99.NA
1.1.6	6)Standardizing terms	1.yes 2.no 0.NA (answerd) 99.NA
1.1.7	7)Stemming Algorithm	1.yes 2.no 0.NA (answerd) 99.NA
1.1.8	8)Term Weighting	1.yes 2.no 0.NA (answerd) 99.NA
1.1.9	9)Phrase identification	1.yes 2.no 0.NA (answerd) 99.NA
1.1.91	-Kind of phrase	0.NA (answerd) 99.NA
1.1.92	-Method used	1.statistical 2.syntactic 3.other 0.NA (answerd) 99.NA
1.1.10	10)Syntactic Parsing	1.yes 2.no 0.NA (answerd) 99.NA
1.1.11	11)Thesaurus/lexical	1.yes 2.no 0.NA (answerd) 99.NA
1.1.12	12)Word sense disambig	1.yes 2.no 0.NA (answerd) 99.NA
1.1.13	13)Spelling checking	1.yes 2.no 0.NA (answerd) 99.NA
1.1.14	14)Correcting them	1.yes 2.no 0.NA (answerd) 99.NA
1.1.15	15)Proper noun identification	1.yes 2.no 0.NA (answerd) 99.NA
1.1.16	16)Method Tokenizing	1.morphological analysis 2.maximal word matching 3.individual characters 4.suffix array 5.other 0.NA (answerd) 99.NA
1.1.17	17)`yomi` in Japanese	1.yes 2.no 0.NA (answerd) 99.NA
1.1.18	18)Method generate `yomi`	0.NA (answerd) 99.NA
1.1.19	19)Other Method Indexing	0.NA (answerd) 99.NA
	1.2.Index data NTCIR-1	
1.2.1	1)Kind of index structures	1.inverted index 2.clusters 3.sign files 4.pat-tree 5.knowledge bases 6.other 0.NA (answerd) 99.NA
	2)Summary of index	
1.2.21	-RUN-ID	
1.2.22	-Total storage (MB)	0.NA (answerd) 99.NA
1.2.23	-Total time build (hour)	0.NA (answerd) 99.NA
1.2.24	-Automatic process (hour)	1.yes 2.no 0.NA (answerd) 99.NA
1.2.25	-Use positional info	1.yes 2.no 0.NA (answerd) 99.NA
	1.3.Data other NTCIR-1	
	1)Internally auxiliary file	

1.3.11	-Domain	0.NA (answerd) 99.NA
1.3.12	-Type of file	1.thesaurus 2.knowledge base 3.lexicon 4.other 0.NA (answerd) 99.NA
1.3.13	-Total storage (MB)	0.NA (answerd) 99.NA
1.3.14	-Number Concept represent	0.NA (answerd) 99.NA
1.3.15	-Type of representation	0.NA (answerd) 99.NA
1.3.16	-Total computer time (hour)	0.NA (answerd) 99.NA
1.3.17	-Total manual time (hour)	0.NA (answerd) 99.NA
1.3.18	-Total m time NTCIR (hour)	0.NA (answerd) 99.NA
1.3.19	-Use of manual labor	1.yes 2.no 0.NA (answerd) 99.NA
1.3.2	2)Externally auxiliary file	1.EDR 2.Chasen morphological dictionary 0.NA (answerd) 99.NA
	2.Query construction	
	2.1.Automatically queries	
2.1.1	1)Time to build query (sec)	0.NA (answerd) 99.NA
	2)Method build query	
2.1.21	-Tokenizing	1.uni-gram 2.bi-gram 3.other n-gram 4.word 5.phrase 6.other 0.NA (answerd) 99.NA
2.1.22	-Phrase extraction	1.morphological analysis 2.other 0.NA (answerd) 99.NA
2.1.23	-Syntactic parsing	1.yes 2.no 0.NA (answerd) 99.NA
2.1.24	-Word sense disambiguation	1.yes 2.no 0.NA (answerd) 99.NA
2.1.25	-Proper noun identification	1.yes 2.no 0.NA (answerd) 99.NA
2.1.26	-Auto expansion queries	1.preciously-constructed tools 2.automatic relevance 3.other 0.NA (answerd) 99.NA
2.1.27	-Auto addition Boolean	1.yes 2.no 0.NA (answerd) 99.NA
2.1.28	-Other	1.full text
	2.2.Manually queries	
2.2.1	1)Time to build query (min)	0.NA (answerd) 99.NA
	2)Person build query	
2.2.21	-Domain expert	1.yes 2.no 3.one part 0.NA (answerd) 99.NA
2.2.22	-Computer system expert	1.yes 2.no 0.NA (answerd) 99.NA
2.2.23	-Other	0.NA (answerd) 99.NA
	3)Tools building query	
2.2.31	-Word frequency	1.yes 2.no 0.NA (answerd) 99.NA
2.2.32	-Knowledge base	1.yes 2.no 0.NA (answerd) 99.NA
2.2.33	-Other lexical tools	1.yes 2.no 0.NA (answerd) 99.NA
2.2.4	4)Method Query Construct	1.term weight 2.boolean operat 3.proximity operat 4.addition term 5.other 0.NA (answerd) 99.NA
	2.3.Interective queries	
2.3.1	1)Initial query a/m	1.auto 2.manual 0.NA (answerd) 99.NA
	2)Person interactive	

2.3.21	-Domain expert	1.yes 2.no 0.NA (answerd) 99.NA
2.3.22	-Computer system expert	1.yes 2.no 0.NA (answerd) 99.NA
2.3.23	-Other	0.NA (answerd) 99.NA
2.3.3	3)Time interaction (min)	0.NA (answerd) 99.NA
2.3.4	4)determine end of interact	0.NA (answerd) 99.NA
	5)Method Interact	
2.3.51	-Auto term reweighting	1.yes 2.no 0.NA (answerd) 99.NA
2.3.52	-Query expansion	1.all terms added 2.top X terms 3.user selected terms 0.NA (answerd) 99.NA
2.3.53	-Query modification manual	1.yes 2.no 0.NA (answerd) 99.NA
	3.Searching	
	3.1.Search time	
3.1.1	1)Run ID	
3.1.2	2)Time to search (sec)	0.NA (answerd) 99.NA
	3.2.Searching method	
3.2.1	1)Vector space model	1.yes 2.no 0.NA (answerd) 99.NA
3.2.2	2)Probabilistic model	1.yes 2.no 0.NA (answerd) 99.NA
3.2.3	3)Other	0.NA (answerd) 99.NA
	3.3.Factors in ranking	
3.3.1	1)TF (Term Frequency)	1.yes 2.no 0.NA (answerd) 99.NA
3.3.2	2)IDF (Inverse Doc Frequency)	1.yes 2.no 0.NA (answerd) 99.NA
3.3.3	3)Other term weights	0.NA (answerd) 99.NA
3.3.4	4)Semantic closement	1.yes 2.no 0.NA (answerd) 99.NA
3.3.5	5)Positional info in doc	1.yes 2.no 0.NA (answerd) 99.NA
3.3.6	6)Syntactic clues	1.yes 2.no 0.NA (answerd) 99.NA
3.3.7	7)Proximity of terms	1.yes 2.no 0.NA (answerd) 99.NA
3.3.8	8)Document length	1.yes 2.no 0.NA (answerd) 99.NA
3.3.9	9)Other	0.NA (answerd) 99.NA
	3.4.Machine information	
3.4.1	1)Machine type experiment	0.NA (answerd) 99.NA
3.4.2	2)Machine d/s	1.dedicated 2.shared 0.NA (answerd) 99.NA
3.4.3	3)Amount hard disc (MB)	0.NA (answerd) 99.NA
3.4.4	4)Amount of RAM (MB)	0.NA (answerd) 99.NA
3.4.5	5)Clock rate CPU (MHz)	0.NA (answerd) 99.NA
	3.5.Others	
3.5.1	1)Brief description other	0.NA (answerd) 99.NA
3.5.2	2)Other	0.NA (answerd) 99.NA

3.5.31	3)Group J native	1.yes 2.no 0.NA (answerd) 99.NA
3.5.32	understand J	1.yes 2.no 0.NA (answerd) 99.NA

NTCIR Workshop Ad Hoc IR Task

Group's ID	Run ID	Q.No.	Contents
ANNO	ANNO	1.2.1	トライ構造
		3.1.2	ただし、繰り返し検索を行わなかった41個の課題について
BKYTR	BKJJDCFU	1.1.4	Longest-match against a lexicon was used to segmented Japanese text into words.
		1.1.16	The longest-match against a dictionary is used to segment Japanese text into tokens.
	BKJJBIDS BKJJBIFU	1.1.16	An adjacent pair of kanji or katakana is treated as one token.
	BKJJBIDS	1.1.5	All hiragana characters were discarded in indexing.
	BKJJBIFU	1.1.11	The dictionary in Chasen and a small Japanese/English dictionary
	BKJJDCFU	1.3.15	Flat text file
		1.3.2	Part of our dictionary was derived from the dictionary in Chasen
		3.3.9	Query length, collection length.
CRL	CRL4	2.1.26	je0のデータから獲得した類義語辞書で自動拡張
	CRL3~10	3.2.3	ロバートソン式を簡略化したもの
	CRL12~15	3.3.3	質問文間のIDF
		3.5.2	trec のツールでは正解の記事番号が複数あるとそれだけ分正解にしてしまうようです。評価に用いられる際には注意して下さい。
DOVE	DOVE1~3	1.1.6	2バイト→1バイト変換
		1.1.19	特になし。(形態素解析結果を品詞、字種、ストップワードでフィルタリング)
	DOVE2 DOVE3	2.3.23	参加者
		2.3.4	原則として1フィードバック
JALAB	JALAB	3.5.1	検索の高速性
JSCB	JSCB1~3	2.3.4	時間制限(20分)
		2.3.52	閾値でカットオフ
KCTRG	KCTRG1 KCTRG2	1.1.19	単語間共起頻度、TF・IDF
		3.1.2	プロトタイプ版であるので発表できません。
		3.3.3	単語共起頻度
		3.3.7	文内一律
KLAB	KLAB1 KLAB2	1.1.4	文字列照合
		1.1.16	文字列照合
		1.2.1	キーワードとテストコレクションデータとの対応表
		2.2.32	与えられた検索課題の、タイトル、日本語概念部分を利用した。
		2.2.4	ORのみ利用
		2.3.3	概念検索モデル(キーワードに着目して、同義表現(単語)を利用した検索モデル)
K3200	K32001 k32002	1.1.5	選択なし(全文字列を索引語に)
		1.1.6	バイト文字を対応するバイト文字に変換した。
		1.1.16	字種ごとにn-gramに分割
		1.2.1	B-Tree
		2.1.27	検索単語をOR結合
NTE15	NTE151 NTE152	1.1.4	辞書を用いた極大単語切り出し方法
		1.1.5	ストップワード以外を選択
		1.1.6	ASCIIコードを対応する2byteコードに変換
		2.1.26	語の類似度(タームベクトル間の内積)に基づく拡張
		3.3.3	文書内共起
RMIT	rmit02	1.1.5	all words
		1.1.8	standard tf.idf
R2D2	R2D22,3	1.2.1	語IDと重みを列挙した単純インデクス
	R2D21~4	3.3.9	著者キーワードから抽出した文書関連性に基づく文書ベクトル拡張
	R2D23,4	2.1.27	近接関係を利用している
sstut	sstut3	1.1.4	情報量の高い部分.
		1.1.5	情報量(idf)
		1.1.8	idfを使用した.
		1.1.92	ダイナミックプログラミングによるMaximum scoring pathで決定.
		1.1.14	行なわないが、修正と等価な効果のある手法を使用した.
		1.1.16	分割しないで部分文字列全体を使用した.
		1.2.1	単純文字列のMatching.
		3.2.3	dpマッチングによるあいまい文字検索.
		3.3.1	ただし全部分文字列に対して計算する.
		3.3.9	dpマッチングによる類似度を用いた.
		3.5.1	dpマッチングを使用した方法で、一つのベースラインを示す目的で

			参加した。言語知識を使用しない方法である。
STIX	STIX1 STIX2	1.1.16	形態素解析と分割用辞書
		1.3.11	DATABASE
		1.3.12	関係語辞書、タイトルパターン辞書、表記揺れ辞書、一般語辞書、 単語bigram辞書
		1.3.13	0.006, 0.003, 0.01, 0.002, 1.6
		1.3.14	335, 233, 128, 55, 814
		2.2.4	文末処理による体言止め化
		STIX1	3.2.3
		3.3.3	係受けの重要度、類似度
TGL	tgl1	1.1.16	文字種分割
		2.1.27	抽出した語に対するand
		2.2.23	情報化学部生
		3.2.3	ブール検索
		3.3.1	検索語の総和
		3.5.1	マルチスレッドによる並列検索
UTS	UTS1 UTS2	1.1.1	2文字以上の全ての部分文字列(文字n-gram)
		1.1.2	2文字以上の全ての部分文字列(文字n-gram)
		1.1.3	2文字以上の全ての部分文字列(文字n-gram)
		1.1.4	Suffix Array
		1.1.5	全ての部分文字列を使用
		1.2.1	Suffix Array
		2.1.28	全文
		3.2.3	文書毎に検索要求の最適な分割を計算する
		3.3.3	部分文字列の長さ
1K	1KJE 1KJE3 1KJE_ij 1KJE.n	1.1.6	全角記号→半角、alphabet→半角
		1.1.16	EUCは1文字、その他は空白で
		1.1.19	OFFSET(位置情報)
		1.3.12	1KJE, 1KJE3 - ntc1-jeのキーワードから自動構築した多言語クラスタ(37270 entry)
			1KJE_ij - 市販の情報処理関係辞書4種から作成したクラスタ(20636 entry)
			1KJE.n - なし
2.1.26	1KJE, 1KJE3 - 1.3.12のntc1-jeから構築した多言語クラスタ		
	1KJE_ij - 市販の情報処理関係辞書4種から作成したクラスタ (Reported in Kando & Aizawa, IRAL'98)		

