



MULTI-LINGUAL INFORMATION RETRIEVAL

Tutorial at SIGIR 2004, June 25, 2004 (morning)
27th International Conference on Research and
Development in Information Retrieval
Sheffield, England

Fredric Gey

University of California, Berkeley

www: <http://ucdata.berkeley.edu/gey.html>

Language Distribution for 2.947 Billion Web Pages 2003

Rank	Language	Number pages	Percent of web
1	English	1690901291	57.37
2	German	256997640	8.72
3	French	120860523	2.88
4	Russian	98422529	2.49
5	Japanese	95372822	2.49
6	Chinese(simp)	93228783	2.50
7	Spanish	89429894	2.46
8	Italian	78104597	2.21
9	Korean	66390095	1.92
10	Dutch	64020603	1.90
11	Portuguese	34193853	1.03
12	Czech	31429288	0.969
13	Swedish	28700267	0.895
14	Polish	27864012	0.877
15	Danish	27622742	0.877
16	Chinese(trad)	21421052	0.687

17	Catalan	19245884	0.623
18	Norwegian	14728161	0.480
19	Hungarian	14504776	0.475
20	Finnish	11856905	0.391
21	Slovak	9236475	0.306
22	Turkish	7587437	0.252
23	Greek	5261482	0.175
24	Hebrew	4563942	0.153
25	Romanian	3993211	0.134
26	Arabic	3882050	0.130
27	Thai	3547806	0.119
28	Croatian	3039926	0.102
29	Estonian	2981244	0.100
30	Bulgarian	2728182	0.092
31	Slovenian	2436725	0.082
32	Byelorussian	2266769	0.077
33	Icelandic	2071663	0.070
34	Lithuanian	1724486	0.058
35	Indonesian	1415089	0.048
	TOTAL	2947327215	

World's Top 20 Languages by Population Speaking

Rank	Language	Country	Population
1	CHINESE MANDARIN [CHN]	China	885000000
2	SPANISH [SPN]	Spain	332000000
3	ENGLISH [ENG]	United Kingdom	322000000
4	BENGALI [BNG]	Bangladesh	189000000
5	HINDI [HND]	India	182000000
6	PORTUGUESE [POR]	Portugal	170000000
7	RUSSIAN [RUS]	Russia	170000000
8	JAPANESE [JPN]	Japan	125000000
9	GERMAN STANDARD [GER]	Germany	98000000
10	CHINESE WU [WUU]	China	77175000
11	JAVANESE [JAN]	Indonesia Java Bali	75500800
12	KOREAN [KKN]	Korea South	75000000
13	FRENCH [FRN]	France	72000000
14	VIETNAMESE [VIE]	Viet Nam	67662000
15	TELUGU [TCW]	India	66350000
16	CHINESE YUE [YUH]	China	66000000
17	MARATHI [MRT]	India	64783000
18	TAMIL [TCV]	India	63075000
19	TURKISH [TRK]	Turkey	59000000
20	URDU [URD]	Pakistan	58000000

Plutonium

In Arabic
Chinese
Greek
Japanese
Korean
Russian
Tamil

بالتونوم عنصر فلزي

钚

πλουτώνιο

プルトニウム

플루토늄

ПЛУТОНИЙ

புரததுநஜியாம்