Subject: Re: Remove unused languages
Posted by crydev on Mon, 06 Jan 2014 15:20:06 GMT
View Forum Message <> Reply to Message

I did some basic calculations, based on the information I found in the disassembler and what I found in files: Core.t and LangInfo.cpp.

The image below shows a few example strings. They are located in Core.t and I assume they are translations for an error or information message in multiple languages. However, in the disassembler you can see that they are all compiled into the executable. I did not measure this, but an educated guess on Core.t, which is 44 kb (~ 40 kb with margins of non-trivial information) is linked into the executable. If just the english translations are linked into the executable, the size could be reduced by roughly: average string length divided by the number of languages built into U++.

The image below shows the language array in LangInfo.cpp. The big rectangle shows the strings that are linked, which are a hell of a lot and they shouldn't be necessary.

The small rectangle shows the fact that some strings even indicate only very small differences. The compilation loaded into IDA Pro is built in VC10 Optimal with String Pooling enabled. Even if these strings could not be eliminated, it would be nice if they could be concatenated in order to remove the redundancy that is generated.

If you need any more information, please tell me.

Regards,
crydev

File Attachments
1) IDA Strings.PNG, downloaded 941 times

Address	Length	Type	String
.rdata:0054FE2C	0000002E	C	range valido superato (limite inferiore: %`).
.rdata:0054FE5C	00000030	C	geldig gebied overschreden (beneden grens: %`).
.rdata:0054FE8C	00000036	C	przekroczono dopuszczalny zakres (dolna granica: %`).
.rdata:0054FEC4	0000002E	C	interval depÄƒÈ™it (margine inferioarÄƒ: %`).
.rdata:0054FEF8	00000054	C	Ð¿Ñ€Ðµ²Ñ‹Ñˆ ен Ð´Ð¾Ð¿Ñƒ�ÑˆÑ‚Ð¼Ñ‹Ð¹ Ð¿Ñ€ÐµÐ´Ðµл (Ð½Ð¶Ð½Ð¸Ð¹ Ð¸Ð
.rdata:0054FF4C	00000037	C	prekroÄenie povolenÃ©ho rozsahu (dolnÃ hranica: %`).
.rdata:0054FF84	00000032	C	prekoraÄena validna granica (donja granica: %`).
.rdata:0054FFB8	0000002E	C	GeÃ§erli aralÄ±k aÅŸÄ±IdÄ± (alt sÄ±nÄ±r: %`).
.rdata:0054FFE8	0000001F	C	è¶…å‡ºæœ‰æ•ˆÈƒå›¢(ä‹é™™:%`).
.rdata:00550009	0000001E	C	¶…å‡ºæœ‰æ•ˆç¯„åœ(ä‹é™™:%`).
.rdata:00550028	00000028	C	exceeded valid range (upper bound: %`).
.rdata:00550050	0000002B	C	rang vÃ lid excedit (lÃmit superior: %`).
.rdata:0055007C	00000034	C	pÅ™ekroÄenÃ povolenÃ©ho rozsahu (hornÃ mez: %`).
.rdata:005500B0	00000037	C	GÃ¼ltigkeitsbereich Ã¼berschritten (obere Grenze: %`).
.rdata:005500E8	0000002F	C	rango vÃilido excedido (lÃmite superior: %`).
.rdata:00550118	0000002D	C	sallitun alueen ulkopuolella (ylÃ¤raja: %`).
.rdata:00550148	00000041	C	validitÃ© de l'intervalle dÃ©passÃ©e (limite supÃ©rieure : %d`).
.rdata:0055018C	00000033	C	Ã©rvÃ©nyes tartomÃ¡ny fÃ¶lÃ¶tt (felsÅ' hatÃ¡r: %`)
.rdata:005501C0	0000002E	C	range valido superato (limite superiore: %`).
.rdata:005501F0	0000002E	C	geldig gebied overschreden (boven grens: %`).
.rdata:00550220	00000037	C	przekroczono dopuszczalny zakres (gÃ³rna granica: %`).
.rdata:00550258	0000002E	C	interval depÄƒÈ™it (margine superioarÄƒ: %`).
.rdata:00550288	00000056	C	Ð¿Ñ€Ðµ²Ñ‹Ñˆ ен Ð´Ð¾Ð¿Ñƒ�ÑˆÑ‚Ð¼Ñ‹Ð¹ Ð¿Ñ€ÐµÐ´Ðµл (Ð²ÐµÑ€Ñ…Ð½Ð¸
.rdata:005502E0	00000037	C	prekroÄenie povolenÃ©ho rozsahu (hornÃ¡ hranica: %`).
.rdata:00550318	00000033	C	prekoraÄena validna granica (gornja granica: %`).
.rdata:0055034C	0000002F	C	GeÃ§erli aralÄ±k aÅŸÄ±IdÄ± (Ã¼st sÄ±nÄ±r: %`).
.rdata:0055037C	0000001F	C	è¶…å‡ºæœ‰æ•ˆÈƒå›¢(äŠé™™:%`).
.rdata:0055039C	0000001F	C	è¶…å‡ºæœ‰æ•ˆç¯„åœ(äŠé™™:%`).
.rdata:005503BC	00000033	C	Invalid date/time: digit expected, found character
.rdata:005503F0	00000038	C	Data/hora incorrecte: s'espera dÃgit, trobat carÃ cter
.rdata:00550428	00000039	C	ChybnÃ© datum/Äas: oÄekÃ¡vÃ¡na ÄÃslice, nalezen znak
.rdata:00550464	0000003F	C	UngÃ¼ltige(s) Datum/Zeit: Ziffer erwartet, Zeichen vorgefunden
.rdata:005504A4	0000003F	C	Fecha/dÃa incorrecto: se espera dÃgito, encontrado carÃicter
.rdata:005504E4	0000003C	C	Virheellinen pvm/aika: KÃ¤ytÃ¤ numeroita kirjaimien sijaan.
.rdata:00550520	0000003C	C	Date/heure invalide : chiffre attendu, caractÃ¨re dÃ©tectÃ©
.rdata:00550560	00000040	C	Ã©rvÃ©nytelen dÃitum/idÅ': szÃimjegy helyett mÃis karakter Ãill
.rdata:005505A0	00000014	C	Data/ora non valida
.rdata:005505B4	00000038	C	Ongeldige datum/tijd: getal verwacht, karakter gevonden
.rdata:005505EC	00000039	C	Niepoprawna data/czas: oczekiwano cyfry, znaleziono znak
.rdata:00550628	0000003A	C	Data/hora invÃilida: digito esperado, encontrou caractere
.rdata:00550664	0000003C	C	DatÄƒ/orÄƒ invalidÄƒ: se aÈ™teaptÄƒ cifrÄƒ, caracter gÄƒsit
.rdata:005506A0	0000006A	C	Ð½ÐµÐ²Ñ€Ð°²Ð¸Ð»ÑŒÐ½Ñ‹Ðµ Ð´Ð°Ñ‚Ð°/²Ñ€ÐµÐ¼Ñ: Ð¾Ð¶Ð¸Ð´Ð°ÐµÑ‚ÑÑ Ñ†Ð¸
.rdata:0055070C	0000003F	C	NesprÃivny dÃitum/Äas: oÄakÃ¡vanÃ¡ ÄÃslice, nÃ¡jdenÃ½ znak
.rdata:0055074C	0000003D	C	Nevalidan datum/vreme: oÄekuje se cifra, pronaÄ'en karakter
.rdata:0055078C	0000003C	C	GeÃ§ersiz tarih/zaman: sayÄ± olmasÄ± gereken yerde harf var
.rdata:005507C8	00000034	C	æ— æ•ˆæ—¥æœŸ/æ—¶é—´:åªè ƒ'ä½æ—æ•°å—,ä¸è Ÿ½ºå—ç¬¦
.rdata:005507FC	00000034	C	ç„¡æˆ^æ—¥æœŸ/æ™,é—":åªè ½æ—æ•,ä½,ä¸è Ÿ½ºç„å—å…ƒ
.rdata:00550830	00000013	C	Invalid date/time:

's'	.rdata:00551AE8	00000008	C	Bengali\t
's'	.rdata:00551B04	00000012	C	Romansh\tRumantsch\t
's'	.rdata:00551B1C	0000000F	C	Mohawk\tKanien'k
's'	.rdata:00551B34	00000011	C	Irish\tGaeilge\t\b<
's'	.rdata:00551B48	0000002A	C	Inuktitut (Latin)\tInuktitut (Kanatami)\t\b]
's'	.rdata:00551B74	00000007	C	Nepali\t
's'	.rdata:00551B94	00000016	C	Mapudungun\tMapudungun\t
's'	.rdata:00551BB0	0000000E	C	Frisian\tFrysk\t
's'	.rdata:00551BC4	00000007	C	Pashto\t
's'	.rdata:00551BD8	0000000F	C	Luxembourgish\tL
's'	.rdata:00551BE9	0000000D	C	tzebuergesch\t
's'	.rdata:00551BFC	00000012	C	Filipino\tFilipino\t
's'	.rdata:00551C14	00000013	C	Bosnian (Cyrillic)\t
's'	.rdata:00551C3C	00000013	C	Serbian (Cyrillic)\t
's'	.rdata:00551C60	0000001C	C	Sami (Inari)\tsÃ¤mikielÃ¢\t$;
's'	.rdata:00551C7C	00000020	C	Sami (Skolt)\tsÃ¤Ã¤mÂ´Ç©iÃµll\t ;
's'	.rdata:00551C9C	00000010	C	Sami (Southern)\t
's'	.rdata:00551CAE	00000012	C	arjelsaemiengiele\t
's'	.rdata:00551CC4	00000010	C	Sami (Southern)\t
's'	.rdata:00551CD6	00000012	C	arjelsaemiengiele\t
's'	.rdata:00551CEC	00000013	C	Sami (Lule)\tjulevus
's'	.rdata:00551D01	00000009	C	megiella\t
's'	.rdata:00551D10	00000013	C	Sami (Lule)\tjulevus
's'	.rdata:00551D25	00000009	C	megiella\t
's'	.rdata:00551D34	00000016	C	Sami (Northern)\tdavvis
's'	.rdata:00551D4C	00000009	C	megiella\t
's'	.rdata:00551D58	00000024	C	Sami (Northern)\tdavvisÃimegiella\t\b;
's'	.rdata:00551D7C	00000016	C	Sami (Northern)\tdavvis
's'	.rdata:00551D94	00000009	C	megiella\t
's'	.rdata:00551DA0	00000020	C	Northern Sotho\tSesotho sa Leboa\t
's'	.rdata:00551DC4	0000000D	C	Zulu\tisiZulu\t
's'	.rdata:00551DD4	0000000F	C	Xhosa\tisiXhosa\t
's'	.rdata:00551DE8	00000010	C	Tswana\tSetswana\t
's'	.rdata:00551DFC	00000011	C	Quechua\trunasimi\t
's'	.rdata:00551E10	00000014	C	Quechua\trunasimi\t\bk
's'	.rdata:00551E24	00000011	C	Quechua\trunasimi\t
's'	.rdata:00551E38	0000000E	C	Maltese\tMalti\t
's'	.rdata:00551E4C	0000000B	C	Maori\tReo M
's'	.rdata:00551E60	0000000E	C	Welsh\tCymraeg\t
's'	.rdata:00551E74	00000017	C	Serbian (Latin)\tsrpski\t
's'	.rdata:00551E90	00000011	C	Bosnian\tbosanski\t
's'	.rdata:00551EA4	00000012	C	Croatian\thrvatski\t
's'	.rdata:00551EBC	00000014	C	Spanish\tEspaÃ±ol\tP\n
's'	.rdata:00551ED0	00000014	C	Spanish\tEspaÃ±ol\tL\n
's'	.rdata:00551EE4	00000014	C	Spanish\tEspaÃ±ol\tH\n
's'	.rdata:00551EF8	00000014	C	Spanish\tEspaÃ±ol\tD\n
's'	.rdata:00551F0C	00000014	C	Spanish\tEspaÃ±ol\t@\n
's'	.rdata:00551F20	00000007	C	Arabic\t