Subject: Remove unused languages
Posted by crydev on Mon, 06 Jan 2014 11:00:27 GMT
View Forum Message <> Reply to Message

Hello,

When I browse through my executable in the disassembler I see a lot of unused languages being compiled into the application. English is sufficient for me, Dutch, Polish, Swedish, etc. are not necessary and in my situation only waste space in the application.

Is there a way to disable or remove them in a way that they are not linked anymore?

Regards,
crydev

Subject: Re: Remove unused languages
Posted by mirek on Mon, 06 Jan 2014 14:47:09 GMT
View Forum Message <> Reply to Message

crydev wrote on Mon, 06 January 2014 06:00Hello,

When I browse through my executable in the disassembler I see a lot of unused languages being compiled into the application. English is sufficient for me, Dutch, Polish, Swedish, etc. are not necessary and in my situation only waste space in the application.

Is there a way to disable or remove them in a way that they are not linked anymore?

Regards,
crydev

That is an interesting request...

Is it worth it? (Can you check?). If it accounts for more than 50KB for console application, I guess we could consider option or compilation flag to leave only EN-US translations...

Mirek

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.

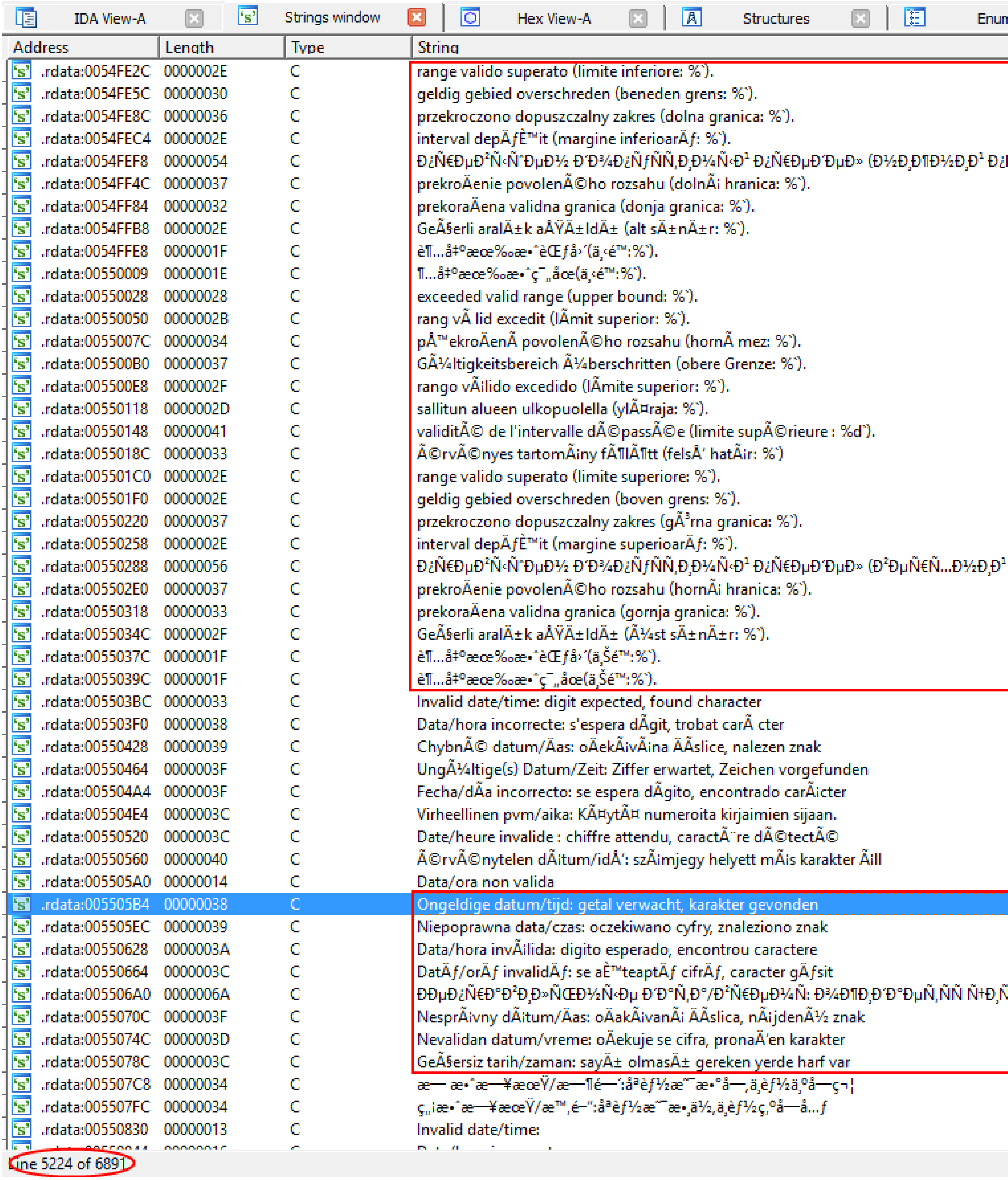
The small rectangle shows the fact that some strings even indicate only very small differences. The compilation loaded into IDA Pro is built in VC10 Optimal with String Pooling enabled. Even if these strings could not be eliminated, it would be nice if they could be concatenated in order to remove the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

If you need any more information, please tell me.

Regards,
crydev

File Attachments

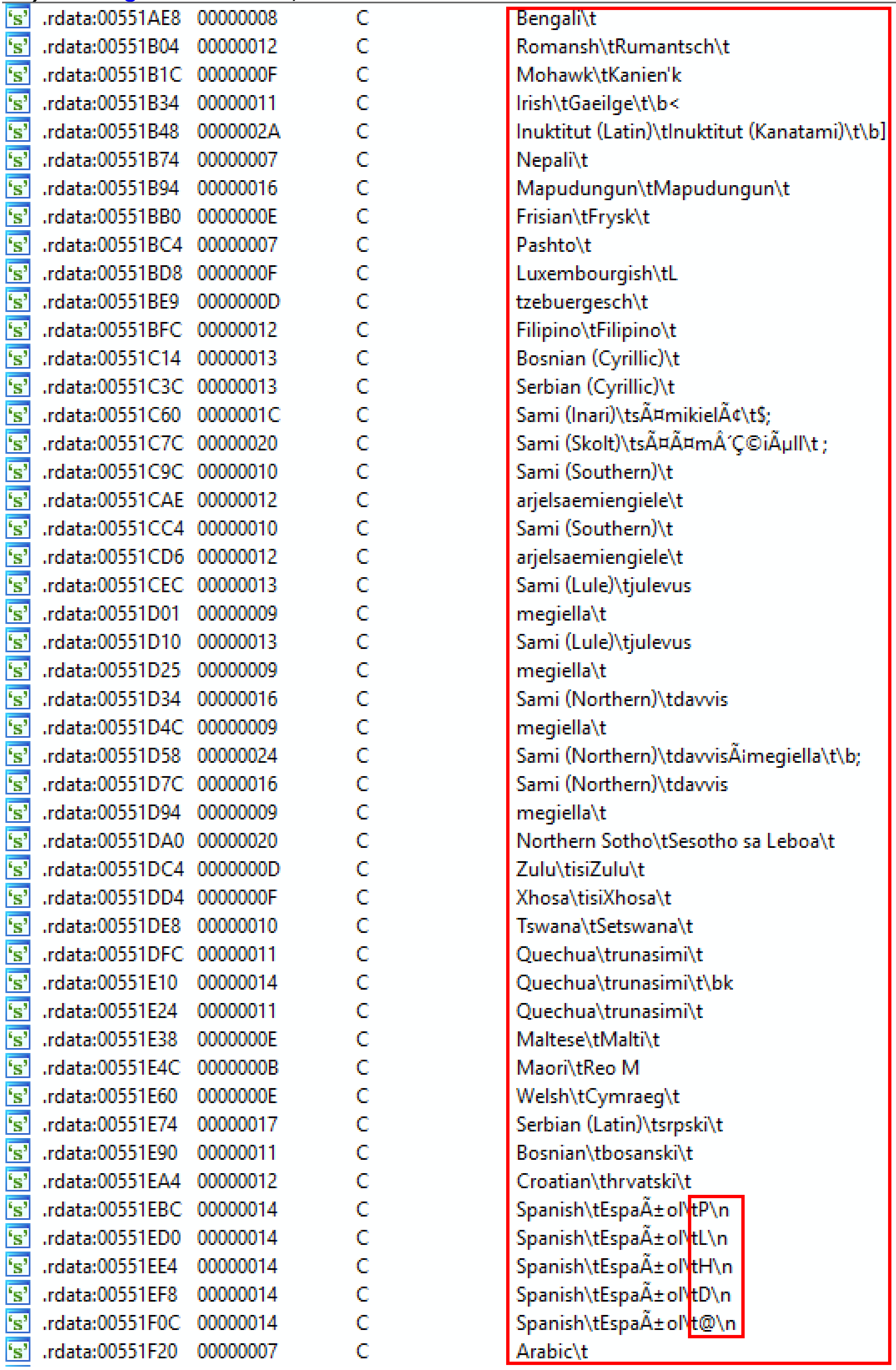
1) IDA Strings.PNG, downloaded 933 times



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 depĂƒÈ™it (margine inferioarĂƒ: %`).
.rdata:0054FEF8	00000054	C	Đ¿Ñ€ĐµĐ²Ñ‹ÑˆĐµĐ½ Đ´Đ¾Đ¿ÑƒÑÑ‚Đ¸Đ¼Ñ‹Đ¹ Đ¿Ñ€ĐµĐ´ĐµĐ» (Đ½Đ¸Đ¶Đ½Đ¸Đ¹ Đ¿Ñ
.rdata:0054FF4C	00000037	C	prekroÄenie povolenĂ©ho rozsahu (dolnĂi hranica: %`).
.rdata:0054FF84	00000032	C	prekoraÄena validna granica (donja granica: %`).
.rdata:0054FFB8	0000002E	C	GeĂ§erli aralÄ±k aÅŸÄ±ldÄ± (alt sÄ±nÄ±r: %`).
.rdata:0054FFE8	0000001F	C	è¶…å‡ºæœ‰æ•ˆèŒƒå›´(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 vĂ lid excedit (lĂmit superior: %`).
.rdata:0055007C	00000034	C	pÅ™ekroÄenĂ povolenĂ©ho rozsahu (hornĂ mez: %`).
.rdata:005500B0	00000037	C	GĂ¼ltigkeitsbereich Ă¼berschritten (obere Grenze: %`).
.rdata:005500E8	0000002F	C	rango vĂilido excedido (lĂmite superior: %`).
.rdata:00550118	0000002D	C	sallitun alueen ulkopuolella (ylĂ¤raja: %`).
.rdata:00550148	00000041	C	validitĂ© de l'intervalle dĂ©passĂ©e (limite supĂ©rieure : %d`).
.rdata:0055018C	00000033	C	Ă©rvĂ©nyes tartomĂiny fĂ¶lĂ¶tt (felsÅ' hatĂir: %`)
.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 (gĂ³rna granica: %`).
.rdata:00550258	0000002E	C	interval depĂƒÈ™it (margine superioarĂƒ: %`).
.rdata:00550288	00000056	C	Đ¿Ñ€ĐµĐ²Ñ‹ÑˆĐµĐ½ Đ´Đ¾Đ¿ÑƒÑÑ‚Đ¸Đ¼Ñ‹Đ¹ Đ¿Ñ€ĐµĐ´ĐµĐ» (Đ²ĐµÑ€Ñ…Đ½Đ¸Đ¹
.rdata:005502E0	00000037	C	prekroÄenie povolenĂ©ho rozsahu (hornĂi hranica: %`).
.rdata:00550318	00000033	C	prekoraÄena validna granica (gornja granica: %`).
.rdata:0055034C	0000002F	C	GeĂ§erli aralÄ±k aÅŸÄ±ldÄ± (Ă¼st sÄ±nÄ±r: %`).
.rdata:0055037C	0000001F	C	è¶…å‡ºæœ‰æ•ˆèŒƒå›´(ä¸Šé™:%`).
.rdata:0055039C	0000001F	C	è¶…å‡ºæœ‰æ•ˆç¯„åœ(ä¸Šé™:%`).
.rdata:005503BC	00000033	C	Invalid date/time: digit expected, found character
.rdata:005503F0	00000038	C	Data/hora incorrecte: s'espera dĂgit, trobat carĂ cter
.rdata:00550428	00000039	C	ChybnĂ© datum/Äas: oÄekĂivĂina ÄĂslice, nalezen znak
.rdata:00550464	0000003F	C	UngĂ¼ltige(s) Datum/Zeit: Ziffer erwartet, Zeichen vorgefunden
.rdata:005504A4	0000003F	C	Fecha/dĂa incorrecto: se espera dĂgito, encontrado carĂicter
.rdata:005504E4	0000003C	C	Virheellinen pvm/aika: KĂ¤ytĂ¤ numeroita kirjaimien sijaan.
.rdata:00550520	0000003C	C	Date/heure invalide : chiffre attendu, caractĂ¨re dĂ©tectĂ©
.rdata:00550560	00000040	C	Ă©rvĂ©nytelen dĂitum/idÅ': szĂimjegy helyett mĂis karakter Ăill
.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 invĂilida: digito esperado, encontrou caractere
.rdata:00550664	0000003C	C	DatĂƒ/orĂƒ invalidĂƒ: se aÈ™teaptĂƒ cifrĂƒ, caracter gĂƒsit
.rdata:005506A0	0000006A	C	ĐĐµĐ¿Ñ€Đ°Đ²Đ¸Đ»ÑŒĐ½Ñ‹Đµ Đ´Đ°Ñ‚Đ°/Đ²Ñ€ĐµĐ¼Ñ: Đ¾Đ¶Đ¸Đ´Đ°ĐµÑ‚ÑÑ Ñ†Đ¸Ñ
.rdata:0055070C	0000003F	C	NesprĂivny dĂitum/Äas: oÄakĂivanĂi ÄĂslica, nĂijdenĂ½ znak
.rdata:0055074C	0000003D	C	Nevalidan datum/vreme: oÄekuje se cifra, pronaÄ'en karakter
.rdata:0055078C	0000003C	C	GeĂ§ersiz tarih/zaman: sayÄ± olmasÄ± gereken yerde harf var
.rdata:005507C8	00000034	C	æ— æ•ˆæ—¥æœŸ/æ—¶é—´:åªèƒ½æ˜¯æ•°å—,ä¸èƒ½ä¸ºå—ç¬¦
.rdata:005507FC	00000034	C	ç„¡æ•ˆæ—¥æœŸ/æ™‚é–“:åªèƒ½æ˜¯æ•¸å—,ä¸èƒ½ç‚ºå—å…ƒ
.rdata:00550830	00000013	C	Invalid date/time:

Line 5224 of 6891

2) Langinfo.PNG, downloaded 988 times



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

Subject: Re: Remove unused languages
Posted by mirek on Mon, 06 Jan 2014 16:45:37 GMT
View Forum Message <> Reply to Message

OK, IMO it makes sense to at least add RM issue about this. Alternatively, we could consider also consider compression (zlib) - but that might be hard(er) to achieve.

Subject: Re: Remove unused languages
Posted by crydev on Mon, 06 Jan 2014 22:24:42 GMT
View Forum Message <> Reply to Message

mirek wrote on Mon, 06 January 2014 17:45OK, IMO it makes sense to at least add RM issue about this. Alternatively, we could consider also consider compression (zlib) - but that might be hard(er) to achieve.

That would be very nice. I have been trying to get my hands on the whereabouts of zlib in U++ GUI applications and why it is being used. I am not sure but I think it has something to do with the iml compression? If so, it has to be researched / tested whether removing zlib actually results in a decrease of image size when images are not compressed anymore and in situations of how many images you should have in order to gain size improvements.

Could you enlighten the purpose of zlib in U++ GUI applications a bit more?

Thanks,
crydev

Subject: Re: Remove unused languages
Posted by mirek on Tue, 07 Jan 2014 08:29:59 GMT
View Forum Message <> Reply to Message

crydev wrote on Mon, 06 January 2014 17:24mirek wrote on Mon, 06 January 2014 17:45OK, IMO it makes sense to at least add RM issue about this. Alternatively, we could consider also consider compression (zlib) - but that might be hard(er) to achieve.

That would be very nice. I have been trying to get my hands on the whereabouts of zlib in U++ GUI applications and why it is being used. I am not sure but I think it has something to do with the iml compression? If so, it has to be researched / tested whether removing zlib actually results in a decrease of image size when images are not compressed anymore and in situations of how many images you should have in order to gain size improvements.

Could you enlighten the purpose of zlib in U++ GUI applications a bit more?

Well, zlib is quite universally used about everywhere fast moderate compression is required. I do not think I will recall all of them, but

.imls are indeed compressed
.brc files can be compressed as well
.tpp files are compressed
HttpRequest has to use zlib to decompress responses
plugin/png is using zlib; png is format of choice for X11 clipboard
Some of RichObject formats use zlib compression
RichEdit spellchecker is using zlib compression
....

Mirek

Subject: Re: Remove unused languages
Posted by Lance on Mon, 27 Dec 2021 01:44:17 GMT
View Forum Message <> Reply to Message

A extremely simply solution to this particular problem is find all *.t that's used by the program, either comment out each of them, or if you don't see any chance you will use them in the future, just delete the content (but not any of the *.t files) of each of them. This works for English only. The saving is not very significant.