Quassel IRC - Feature #14

Handle different encodings and UTF8

10/29/2006 04:36 PM - Sputnick

Status:	Closed	**Start date:**	
Priority:	High	**Due date:**	
Assignee:	Sputnick	**% Done:**	0%
Category:		**Estimated time:**	0.00 hour
Target version:	0.2.0-alpha1		
OS:	Any		

Description

We should have support (automatic) encoding selection for both receiving and sending. Unicode detection is crucial.

History

#1 - 06/18/2007 12:14 PM - Sputnick

I have recently read up on how utf8 works, and gather that utf8 detection should be no real problem. In particular, any string that passes as valid utf8 can be considered being utf8. Chances for a false positive are *very* slim, since this would involve character combinations that should not exist in any language.

So we would check any incoming string for being utf8, and use that if it actually is utf8, and otherwise apply a default/configured decoding.

Encodings for both sending and receiving should be configurable globally (i.e default) as well as per-channel. Also we would need encoding settings for the communication between client and server. This should only affect channelnames though (think chan names with non-ascii chars in them).

#2 - 07/08/2007 04:56 PM - Sputnick

Just a quick update: utf8 detection is done (as a function in util.*) and works as expected. Still need to do the conversion where appropriate, which means some work in server.*, in particular. Or maybe string decoding should even move to the GUI?

#3 - 12/25/2007 02:32 AM - Sputnick

We have the basic infrastructure in place, utf8 detection for receiving messages works. A user UI for configuring encodings is not yet available (we have hardcoded ISO-8859-15 for the time being), but that'll come in due time. Closing this bug, since the crucial issues are taken care of.