package com.twitter.tweetypie.tweettext

/\*\*

\* Code used to convert raw user-provided text into an allowable form.

\*/

object PartialHtmlEncoding {

/\*\*

\* Replaces all `<`, `>`, and '&' chars with "&lt;", "&gt;", and "&amp;", respectively.

\*

\* Tweet text is HTML-encoded at tweet creation time, and is stored and processed in encoded form.

\*/

def encode(text: String): String = {

val buf = new StringBuilder

text.foreach {

case '<' => buf.append("&lt;")

case '>' => buf.append("&gt;")

case '&' => buf.append("&amp;")

case c => buf.append(c)

}

buf.toString

}

private val AmpLtRegex = "&lt;".r

private val AmpGtRegex = "&gt;".r

private val AmpAmpRegex = "&amp;".r

private val partialHtmlDecoder: (String => String) =

((s: String) => AmpLtRegex.replaceAllIn(s, "<"))

.andThen(s => AmpGtRegex.replaceAllIn(s, ">"))

.andThen(s => AmpAmpRegex.replaceAllIn(s, "&"))

/\*\*

\* The opposite of encode, it replaces all "&lt;", "&gt;", and "&amp;" with

\* `<`, `>`, and '&', respectively.

\*/

def decode(text: String): String =

decodeWithModification(text) match {

case Some(mod) => mod.updated

case None => text

}

/\*\*

\* Decodes encoded entities, and returns a `TextModification` if the text was modified.

\*/

def decodeWithModification(text: String): Option[TextModification] =

TextModification.replaceAll(

text,

AmpLtRegex -> "<",

AmpGtRegex -> ">",

AmpAmpRegex -> "&"

)

}