package com.twitter.tweetypie

package store

import com.twitter.tweetypie.thriftscala.\_

import scala.util.matching.Regex

object MediaIndexHelper {

/\*\*

\* Which tweets should we treat as "media" tweets?

\*

\* Any tweet that is not a retweet and any of:

\* - Is explicitly marked as a media tweet.

\* - Has a media entity.

\* - Includes a partner media URL.

\*/

def apply(partnerMediaRegexes: Seq[Regex]): Tweet => Boolean = {

val isPartnerUrl = partnerUrlMatcher(partnerMediaRegexes)

tweet =>

getShare(tweet).isEmpty &&

(hasMediaFlagSet(tweet) ||

getMedia(tweet).nonEmpty ||

getUrls(tweet).exists(isPartnerUrl))

}

def partnerUrlMatcher(partnerMediaRegexes: Seq[Regex]): UrlEntity => Boolean =

\_.expanded.exists { expandedUrl =>

partnerMediaRegexes.exists(\_.findFirstIn(expandedUrl).isDefined)

}

def hasMediaFlagSet(tweet: Tweet): Boolean =

tweet.coreData.flatMap(\_.hasMedia).getOrElse(false)

}