package com.twitter.timelines.prediction.features.escherbird

import com.twitter.tweetypie.thriftscala.Tweet

import scala.collection.JavaConverters.\_

object EscherbirdFeaturesConverter {

val DeprecatedOrTestDomains = Set(1L, 5L, 7L, 9L, 14L, 19L, 20L, 31L)

def fromTweet(tweet: Tweet): Option[EscherbirdFeatures] = tweet.escherbirdEntityAnnotations.map {

escherbirdEntityAnnotations =>

val annotations = escherbirdEntityAnnotations.entityAnnotations

.filterNot(annotation => DeprecatedOrTestDomains.contains(annotation.domainId))

val tweetGroupIds = annotations.map(\_.groupId.toString).toSet.asJava

val tweetDomainIds = annotations.map(\_.domainId.toString).toSet.asJava

// An entity is only unique within a given domain

val tweetEntityIds = annotations.map(a => s"${a.domainId}.${a.entityId}").toSet.asJava

EscherbirdFeatures(tweet.id, tweetGroupIds, tweetDomainIds, tweetEntityIds)

}

}