package com.twitter.search.ingester.pipeline.twitter;

import org.apache.commons.pipeline.StageException;

import org.apache.commons.pipeline.validation.ConsumedTypes;

import org.apache.commons.pipeline.validation.ProducesConsumed;

import com.twitter.search.common.relevance.classifiers.TweetOffensiveEvaluator;

import com.twitter.search.common.relevance.entities.TwitterMessage;

import com.twitter.search.common.relevance.scorers.TweetTextScorer;

import com.twitter.search.common.relevance.text.TweetParser;

import com.twitter.search.ingester.model.IngesterTwitterMessage;

@ConsumedTypes(TwitterMessage.class)

@ProducesConsumed

public class TextUrlsFeatureExtractionStage extends TwitterBaseStage

<IngesterTwitterMessage, IngesterTwitterMessage> {

private final TweetParser tweetParser = new TweetParser();

private TweetOffensiveEvaluator offensiveEvaluator;

private final TweetTextScorer tweetTextScorer = new TweetTextScorer(null);

@Override

protected void doInnerPreprocess() {

innerSetup();

}

@Override

protected void innerSetup() {

offensiveEvaluator = wireModule.getTweetOffensiveEvaluator();

}

@Override

public void innerProcess(Object obj) throws StageException {

if (!(obj instanceof IngesterTwitterMessage)) {

throw new StageException(this, "Object is not a TwitterMessage instance: " + obj);

}

IngesterTwitterMessage message = IngesterTwitterMessage.class.cast(obj);

extract(message);

emitAndCount(message);

}

private void extract(IngesterTwitterMessage message) {

tweetParser.parseUrls(message);

offensiveEvaluator.evaluate(message);

tweetTextScorer.scoreTweet(message);

}

@Override

protected IngesterTwitterMessage innerRunStageV2(IngesterTwitterMessage message) {

extract(message);

return message;

}

}