CMU 11791 fall 2013 Homework 3

Wei Zhang

Id: weizhan1

Task1.2:

1. Collection reader uses the FileSystemCollectionReader as directed
2. A Cas Consumer is created to output the evaluation in the console.
3. CPE descriptor is named as hw3-iD-CPE.xml, which is an aggregation of collection reader, hw2-weizhan1-aae, and cas consumer.

Task 2.2:
1. created scnlp-weizhan1-client.xml for ClearTK toolkit.
2. The client is added in the CPE before the hw2-weizhan1-aae.xml, so that the GenAnswerScore.java, which is the AnswerScore generation model could use the annotations in the CAS that are generated by the scnlp client.
   Note that the NamedEntityMention is added to the final GenAnswerScore.java, and created a GenAnswerScoreUsingNE.java. Correspondingly, a new aae is created for the new pipeline, with name hw3-weizhan1-aae-WithClearTK.xml. CPE is tested with this WithClearTK version of the hw2 aae.
3. The CPE is named as hw3-weizhan1-CPE.xml.
4. The way that NamedEntityMention are used is as a "bag of Named Entity Mentions". After adding this feature, the P@N does not change, because the P@N for homework2 has already been 100%. Adding NER annotations won't hurt, though.
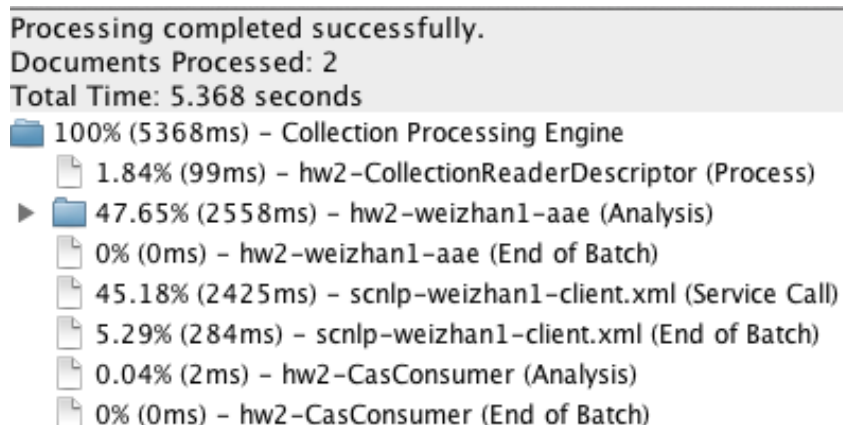5. Speed Comparison:

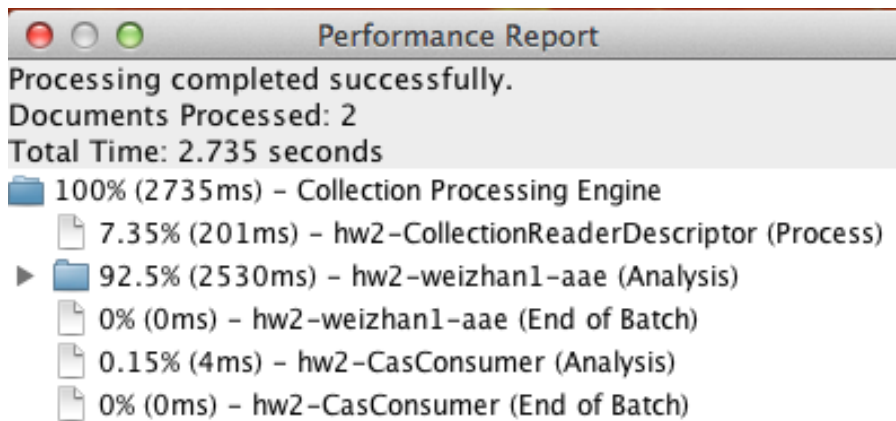Fig.1. The efficnency breakdown after using the scnlp client

Fig.2. The efficnency breakdown before using the scnlp client

As we can see from the comparison of Fig.1 and Fig.2, adding the scnlp remote call will increase the total run time from 2.7 sec to 5.3 sec. This test is done in a network environment outside of cmu campus. The initialization time for the local version is 2.5 sec, which is is comparable to the network call time, which is around 2.4 sec. There is no advantage to call the network service in this setting.

Task 2.3
1. The deployment of the hw2 aae is created with name hw2-weizhan1-aae-deploy.xml
2. Broker started succesfully.
3. Client for hw2-weizhan1-aae-client.xml is created.
4. A CPE called hw2-weizhan1-aae-as-CPE.xml is created to test the service.

Task 2.4
1. The Stanford CoreNLP annotator is set up locally, and the comparison can be sen in Fig.1 and Fig.2. As is mentioned above, there is no advantage of the remote access because of the out of campus network. The corenlp setup can be seen in the WordFeatureAnnotator.java in the annotators folder.
2. The POS-tagging, lemma, and parsing are incorporated.

Task 2.5
Done.