

TestPlan1.5.2

Test Plan for Apache OpenNLP 1.5.2

This page contains the test plan for the 1.5.2 release.

The 1.5.2 release does not introduce any changes to the feature generation expect for the name finder which might generate different token class features for words with special letters.

Compatibility Test with OpenNLP 1.5.0 SourceForge Models

The 1.5.0 SourceForge models must be fully compatible with the 1.5.2 release. In this test all the English models are tested for compatibility on the English 300k sentences Leipzig Corpus. It is tested that the output produced with the same model by both versions has the same md5 hash.

Component	Model	Perf 1.5.1	Perf 1.5.2	Tester	Passed	Comment
Sentence Detector	en-sent.bin	42186.7 sent /s		joern	no	It did not pass because of OPENNLP-202. The diff showed that in the first 20 compared cases didn't made a mistake compared to 1.5.1.
Tokenizer	en-token.bin	3091.8 sent/s	2300.4 sent/s	joern	yes	
Name Finder	en-ner-person.bin	614.4 sent/s	650.6 sent/s	joern	yes	output identical, measurement was done on a idle system, the new name finder is roughly 10% faster
POS Tagger	en-pos-maxent.bin	732.1 sent/s	816.9 sent/s	joern	yes	
POS Tagger	en-pos-perceptron.bin	1110.6 sent/s		joern	no	Perceptron normalization was changed.
Chunker	en-chunker.bin	167,3 sent/s	166.4 sent/s	joern	yes	
Parser	en-parser-chunking.bin	11.6 sent/s		joern	no	A very few sentences are parsed differently due to OPENNLP-233. The parser code itself it not affected by this only the code in the cmd line package.

Note: Test was done on MacBook Pro 13" 7.1, 2.66 GHz Core 2 Duo, 8GB Ram, 256GB SSD running OS X 10.6.6 and Java 1.6.0_26 64-Bit Server. The performance varies because light weight tasks have been performed in the background while testing.

Note: "Concurrent" in the comment means that both tests where started at the same time.

Regression Test Training with (private) English data

The training of both versions with the same data must produce a model with identical output. The model output is tested with the procedure from the previous test.

To pass the test the event hash and the model output must be identical.

Component	Model	Training Time 1.5.1	Training Time 1.5.2	Tester	Passed	Comment
Sentence Detector	en-sent.bin	0m11.255s		joern	no	The new version is more accurate due to OPENLP-202.
Tokenizer	en-token.bin	2m30.115s	1m35.414s	joern	yes	
POS Tagger	en-pos-maxent.bin			joern	yes	Test is still done, because tagdict is not tested with public data
POS Tagger	en-pos-perceptron.bin			joern	no	Perceptron code was changed
Parser	en-parser-chunking.bin	138m9.045s		joern	no	There are small differences due to OPENNLP-233. Changes to the training code seems not to cause regressions.

Note: Time was measured with the time command, the value is the "real" time value.

Performance test with public data

Test the tagging performance with all the publicly available training and test data for various languages.

It is assumed that the training will be done with a cutoff of 5 and 100 iterations, if different values are used please write them into the comment.

Component	Data	Tester	Tagging Perf 1.5.1	Tagging Perf 1.5.2	Comment
Sentence Detector					
Tokenizer					
Name Finder	CONLL 2002 Dutch Person ned.testa	jkosin	Precision: 0.7906976744186046 Recall: 0.48364153627311524 F-Measure: 0.6001765225066196	Precision: 0.7552941176470588 Recall: 0.4566145092460882 F-Measure: 0.5691489361702128	Performance Change due to OPENNLP-294 and more...

Name Finder	CONLL 2002 Dutch Person ned.testb	josin	Precision: 0.8527980535279805 Recall: 0.6384335154826958 F-Measure: 0.7302083333333333	Precision: 0.8505025125628141 Recall: 0.6165755919854281 F-Measure: 0.7148891235480465	
Name Finder	CONLL 2002 Dutch Organization ned.testa	josin	Precision: 0.8386075949367089 Recall: 0.38629737609329445 F-Measure: 0.5289421157684631	Precision: 0.8561872909698997 Recall: 0.37317784256559766 F-Measure: 0.5197969543147207	
Name Finder	CONLL 2002 Dutch Organization ned.testb	josin	Precision: 0.7784200385356455 Recall: 0.4580498866213152 F-Measure: 0.5767309064953604	Precision: 0.7830374753451677 Recall: 0.4501133786848073 F-Measure: 0.5716342692584593	
Name Finder	CONLL 2002 Dutch Location ned.testa	josin	Precision: 0.8362831858407079 Recall: 0.3945720250521921 F-Measure: 0.5361702127659574	Precision: 0.8458333333333333 Recall: 0.42379958246346555 F-Measure: 0.564673157162726	
Name Finder	CONLL 2002 Dutch Location ned.testb	josin	Precision: 0.854251012145749 Recall: 0.545219638248941 F-Measure: 0.665615141955836	Precision: 0.8816326530612245 Recall: 0.5581395348637209 F-Measure: 0.6835443037974683	
Name Finder	CONLL 2002 Dutch Misc ned.testa	josin	Precision: 0.8300492610837439 Recall: 0.4505347593582888 F-Measure: 0.5840554592720971	Precision: 0.8354114713216958 Recall: 0.44786096256684493 F-Measure: 0.5831157528285466	
Name Finder	CONLL 2002 Dutch Misc ned.testb	josin	Precision: 0.8373205741626795 Recall: 0.44229149115417016 F-Measure: 0.5788313120176405	Precision: 0.8264984227129337 Recall: 0.44144903117101936 F-Measure: 0.5755079626578803	
Name Finder	CONLL 2002 Combined ned.testa	josin	Precision: 0.7906976744186046 Recall: 0.48364153627311524 F-Measure: 0.6001765225066196	Precision: 0.6509695290858726 Recall: 0.628822629969419 F-Measure: 0.6397044526540929	1000 iterations OPENNLP-335 Exporting of all tags...
Name Finder	CONLL 2002 Dutch Combined ned.testb	josin	Precision: 0.8527980535279805 Recall: 0.6384335154826958 F-Measure: 0.7302083333333333	Precision: 0.6869929337869668 Recall: 0.6660746003552398 F-Measure: 0.6763720690543674	1000 iterations
Name Finder	CONLL 2002 Spanish Person esp.testa	josin	Precision: 0.8982630272952854 Recall: 0.5924713584288053 F-Measure: 0.7140039447731755	Precision: 0.9010695187165776 Recall: 0.5515548281505729 F-Measure: 0.684263959390863	
Name Finder	CONLL 2002 Spanish Person esp.testb	josin	Precision: 0.9008 Recall: 0.7659863945578231 F-Measure: 0.8279411764705882	Precision: 0.9195205479452054 Recall: 0.7306122448979592 F-Measure: 0.8142532221379833	
Name Finder	CONLL 2002 Spanish Organization esp.testa	josin	Precision: 0.8216258879242304 Recall: 0.6123529411764705 F-Measure: 0.7017189079878665	Precision: 0.8288942695722357 Recall: 0.6041176470588235 F-Measure: 0.6988771691051379	
Name Finder	CONLL 2002 Spanish Organization esp.testb	josin	Precision: 0.8009331259720062 Recall: 0.7357142857142858 F-Measure: 0.7669396872673119	Precision: 0.8036277602523659 Recall: 0.7278571428571429 F-Measure: 0.7638680659670164	
Name Finder	CONLL 2002 Spanish Location esp.testa	josin	Precision: 0.7481789802289281 Recall: 0.7306910569105691 F-Measure: 0.739331619537275	Precision: 0.7743016759776536 Recall: 0.7042682926829268 F-Measure: 0.7376263970196913	
Name Finder	CONLL 2002 Spanish Location esp.testb	josin	Precision: 0.8226221079691517 Recall: 0.5904059040590406 F-Measure: 0.6874328678839956	Precision: 0.8301886792452831 Recall: 0.5682656826568265 F-Measure: 0.6746987951807228	
Name Finder	CONLL 2002 Spanish Misc esp.testa	josin	Precision: 0.6446886446886447 Recall: 0.3955056179775281 F-Measure: 0.49025069637883006	Precision: 0.6492890995260664 Recall: 0.30786516853932583 F-Measure: 0.4176829268292683	
Name Finder	CONLL 2002 Spanish Misc esp.testb	josin	Precision: 0.6595744680851063 Recall: 0.36578171091445427 F-Measure: 0.4705882352941176	Precision: 0.686046511627907 Recall: 0.3480825958702065 F-Measure: 0.461839530332681	
Name Finder	CONLL 2002 Spanish Combined esp.testa	josin	Precision: 0.8982630272952854 Recall: 0.5924713584288053 F-Measure: 0.7140039447731755	Precision: 0.7005423249233671 Recall: 0.6828315329809239 F-Measure: 0.6915735567970205	1000 iterations
Name Finder	CONLL 2002 Spanish Combined esp.testb	josin	Precision: 0.9008 Recall: 0.7659863945578231 F-Measure: 0.8279411764705882	Precision: 0.756635931824532 Recall: 0.7611017425519955 F-Measure: 0.7588622670589884	1000 iterations
Name Finder	CONLL 2003 English Person eng.testa	josin	Precision: 0.9352201257861635 Recall: 0.8072747014115093 F-Measure: 0.8665501165501166	Precision: 0.9523195876288659 Recall: 0.8023887079261672 F-Measure: 0.8709487330583382	
Name Finder	CONLL 2003 English Person eng.testb	josin	Precision: 0.8873546511627907 Recall: 0.7551020408163265 F-Measure: 0.8159037754761109	Precision: 0.9391727493917275 Recall: 0.7161410018552876 F-Measure: 0.8126315789473685	
Name Finder	CONLL 2003 English Organization eng.testa	josin	Precision: 0.8528584817244611 Recall: 0.6785980611483967 F-Measure: 0.7558139534883722	Precision: 0.8768046198267565 Recall: 0.6793437733035048 F-Measure: 0.7655462184873949	
Name Finder	CONLL 2003 English Organization eng.testb	josin	Precision: 0.8263422818791947 Recall: 0.5930162552679109 F-Measure: 0.6905012267788293	Precision: 0.8435980551053485 Recall: 0.6267308850090307 F-Measure: 0.7191709844559586	
Name Finder	CONLL 2003 English Location eng.testa	josin	Precision: 0.9283837056504599 Recall: 0.769188894937398 F-Measure: 0.8413218219708247	Precision: 0.9361421988150099 Recall: 0.7740881872618399 F-Measure: 0.8474374255065554	
Name Finder	CONLL 2003 English Location eng.testb	josin	Precision: 0.9156180606957809 Recall: 0.7416067146282974 F-Measure: 0.8194766478966545	Precision: 0.9206349206349206 Recall: 0.7302158273381295 F-Measure: 0.8144433299899699	
Name Finder	CONLL 2003 English Misc eng.testa	josin	Precision: 0.8539007092198582 Recall: 0.6529284164859002 F-Measure: 0.7400122925629993	Precision: 0.9027982326951399 Recall: 0.6648590021691974 F-Measure: 0.7657173928794503	
Name Finder	CONLL 2003 English Misc eng.testb	josin	Precision: 0.8599137931034483 Recall: 0.5683760683760684 F-Measure: 0.6843918086174958	Precision: 0.8592436974789915 Recall: 0.5826210826210826 F-Measure: 0.694397283514092	
Name Finder	CONLL 2003 English Combined eng.testa	josin	Precision: 0.8601818493738206 Recall: 0.8438236284079434 F-Measure: 0.8519242205420101	Precision: 0.861812521618817 Recall: 0.8386065297879501 F-Measure: 0.8500511770726714	1000 iterations
Name Finder	CONLL 2003 English Combined eng.testb	josin	Precision: 0.8036415565869333 Recall: 0.7970963172804533 F-Measure: 0.8003555555555556	Precision: 0.8041311831853597 Recall: 0.7857648725212465 F-Measure: 0.7948419450165667	1000 iterations
Name Finder	CONLL 2003 German Person deu.testa	joern	Precision: 0.8602620087336245 Recall: 0.28122769450392576 F-Measure: 0.4238838084991931	Precision: 0.9132653061224489 Recall: 0.25553176302640973 F-Measure: 0.3993307306190742	

Name Finder	CONLL 2003 German Person deu.testb	joern	Precision: 0.878 Recall: 0.3673640167364017 F-Measure: 0.5179941002949853	Precision: 0.8732106339468303 Recall: 0.3573221757322176 F-Measure: 0.507125890736342	
Name Finder	CONLL 2003 German Organization deu. testa	joern	Precision: 0.8365695792880259 Recall: 0.41659951651893634 F-Measure: 0.5562130177514794	Precision: 0.8407224958949097 Recall: 0.4125705076551168 F-Measure: 0.5535135135135135	
Name Finder	CONLL 2003 German Organization deu. testb	joern	Precision: 0.7942583732057417 Recall: 0.4294954721862872 F-Measure: 0.5575146935348446	Precision: 0.8014705882352942 Recall: 0.4230271668822768 F-Measure: 0.5537679932260795	
Name Finder	CONLL 2003 German Location deu.testa	joern	Precision: 0.7362637362637363 Recall: 0.34038950042337 F-Measure: 0.4655471916618414	Precision: 0.7816326530612245 Recall: 0.32430143945808637 F-Measure: 0.45840813883901854	
Name Finder	CONLL 2003 German Location deu.testb	joern	Precision: 0.75 Recall: 0.3652173913043478 F-Measure: 0.4912280701754385	Precision: 0.8033826638477801 Recall: 0.3671497584541063 F-Measure: 0.5039787798408487	
Name Finder	CONLL 2003 German Misc deu.testa	joern	Precision: 0.7213930348258707 Recall: 0.14356435643564355 F-Measure: 0.2394715111478117	Precision: 0.7055555555555556 Recall: 0.12574257425742574 F-Measure: 0.21344537815126052	
Name Finder	CONLL 2003 German Misc deu.testb	joern	Precision: 0.6198830409356725 Recall: 0.1582089552238806 F-Measure: 0.2520808561236623	Precision: 0.6601307189542484 Recall: 0.15074626865671642 F-Measure: 0.2454434993924666	
Name Finder	CONLL 2003 German Combined deu.testa	joern	Precision: 0.7675205413243112 Recall: 0.32857438444030623 F-Measure: 0.46015647638365687	Precision: 0.7718859429714857 Recall: 0.319263397475688 F-Measure: 0.4516978922716628	
Name Finder	CONLL 2003 German Combined deu.testb	joern	Precision: 0.7553418803418803 Recall: 0.3849714130138851 F-Measure: 0.5100090171325519	Precision: 0.7467566165023353 Recall: 0.3917778382793357 F-Measure: 0.5139285714285715	
POS Tagger	CONLL 2006 Danish	joern	Accuracy: 0.9511278195488722	Accuracy: 0.9511278195488722	
POS Tagger	CONLL 2006 Dutch	joern	Accuracy: 0.9324977618621307	Accuracy: 0.9324977618621307	
POS Tagger	CONLL 2006 Portuguese	joern	Accuracy: 0.9659110277825124	Accuracy: 0.9659110277825124	
POS Tagger	CONLL 2006 Swedish	joern	Accuracy: 0.9275106082036775	Accuracy: 0.9275106082036775	
Chunker	CONLL 2000	colen	Precision: 0.9255923572240226 Recall: 0.9220610430991112 F-Measure: 0.9238233255623465	Precision: 0.9257575757575758 Recall: 0.9221868187154117 F-Measure: 0.9239687473746113	Perf change due to OPENNLP-242
Chunker	Arvores Deitadas (10-fold cross-validation)	colen	Precision: 0.9413606010016694 Recall: 0.9379938451301671 F-Measure: 0.9396742073907428	Precision: 0.9403445830378374 Recall: 0.9373141775994345 F-Measure: 0.9388269348910339	Perf change due to OPENNLP-242 and OPENNLP-186

The results of the tagging performance might differ compared to the 1.5.0 release since a precision bug in the calculation of the score has been fixed: <https://issues.apache.org/jira/browse/OPENNLP-59>
A problem was corrected for the CoNLL 02 data being improperly converted to the wrong encoding.

Test UIMA Integration

The test ensures that the Analysis Engine can run and not not crash trough simple runtime time code errors. We need to add more sophisticated testing with the next releases.

Analysis Engine	Tester	Passed	Comment
Sentence Detector	joern	yes	
Sentence Detector Trainer	joern	yes	Trained with a UIMA pipeline
Tokenizer ME	joern	yes	
Tokenizer Trainer	joern	yes	Trained with a UIMA pipeline
Name Finder	joern	yes	
Name Finder Trainer	joern	yes	Trained with a UIMA pipeline
Chunker	joern	yes	as part of sample pear
Chunker Trainer			
POS Tagger	joern	yes	as part of sample pear
POS Tagger Trainer			Trained and tested with cmd line tool
Parser			
createPear.sh	joern	yes	
Sample PEAR	joern	yes	installed and run over sample text

Distribution Review

Please ensure that the listed files below are included in the distributions and are in a good state.

Package	File or Test	Tester	Passed	Comment
Binary	LICENSE	joern	yes	AL 2.0 and BSD for JWNL
Binary	NOTICE	joern	yes	standard notice, dates are correct. JWNL is mentioned
Binary	README	colen, jason, james, joern	yes	File was reviewed on the dev list.
Binary	RELEASE_NOTES.html	joern, james	yes	issue list is generated correctly
Binary	Test signatures: .md5, .sha1, .asc	joern	yes	rc4
Binary	JIRA issue list created	joern	yes	
Binary	Contains maxent, tools, uima and jwnl jars	joern	yes	
Source	LICENSE	joern	yes	standard AL 2.0 file
Source	NOTICE	joern	yes	standard notice, dates are correct
Source	Test signatures: .md5, .sha1, .asc	joern	yes	rc4
Source	Can build from source?	joern	yes	Test should be done without jwnl and opennlp in local m2 repo. Test was done on Ubuntu 10.10.