

TestPlan1.6.0

Test Plan for Apache OpenNLP 1.6.0

This page contains the test plan for the 1.6.0 release.

The 1.6.0 release introduces API changes.

Apache OpenNLP 1.6.0 requires Java 1.7.

Compatibility Test with OpenNLP 1.5.0 SourceForge Models

The 1.5.0 SourceForge models must be fully compatible with the 1.6.0 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.

Load: the time in seconds to Apache OpenNLP took to load the model.

Avg: the average throughput of the module, in sentences per second.

| Component | Model | Perf 1.5.2 | Perf 1.5.3 | Perf 1.6.0 | Tester | Passed | Comment |
|-------------------|------------------------|-------------------------------------|-------------------------------------|-------------------------------------|---------|--------|--|
| Sentence Detector | en-sent.bin | load: 0,093s avg: 34945,0 sent/s | load: 0,099s avg: 75960,9 sent/s | load: 0,106s avg: 76133,7 sent/s | William | False | Output does not match. |
| Tokenizer | en-token.bin | load: 0,234s avg: 3721,2 sent/s | load: 0,201s avg: 6108,4 sent/s | load: 0,198s avg: 6125,9 sent/s | William | Yes | RC2 |
| Name Finder | en-ner-person.bin | load: 1,133s avg: 1451,2 sent/s | load: 1,152s avg: 1615,6 sent/s | load: 1,129s avg: 1559,2 sent/s | William | Yes | RC2 |
| POS Tagger | en-pos-maxent.bin | load: 1,764s avg: 1218,9 sent/s | load: 1,703s avg: 1469,8 sent/s | load: 1,435s avg: 1431,2 sent/s | William | Yes | RC2 |
| POS Tagger | en-pos-perceptron.bin | load: 1,309s avg: 1839,0 sent/s | load: 1,124s avg: 2210,5 sent/s | load: 1,170s avg: 2202,7 sent/s | William | Yes | RC2 |
| Chunker | en-chunker.bin | load: 0,667s avg: 439,5 sent/s | load: 0,510s avg: 446,7 sent/s | load: 0,562s avg: 1332,5 sent/s | Jörn | Yes | RC4. Perf was not updated. Should be ok. |
| Parser | en-parser-chunking.bin | load: 8,353s avg: 27,4 sent/s | load: 7,335s avg: 33,2 sent/s | load: 7,200s avg: 44,3 sent/s | Jörn | Yes | RC4. Perf was not updated. Should be ok. |

Note: Test was done on MacBook Pro 15", 2 GHz Core i7, 16GB Ram, 500GB HD running OS X 10.10.1 and Java 1.7.0_71-b14. The performance varies because light weight tasks have been performed in the background while testing.

Note: "Concurrent" in the comment means that both tests were 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 | Tester | Passed | Comment |
|-------------------|------------------------|--------|--------|---|
| Sentence Detector | en-sent.bin | Jörn | Yes | RC2 |
| Tokenizer | en-token.bin | Jörn | Yes | RC2 |
| POS Tagger | en-pos-maxent.bin | Jörn | Yes | RC2 |
| POS Tagger | en-pos-perceptron.bin | Jörn | Yes | RC2 |
| Parser | en-parser-chunking.bin | Jörn | Yes | RC4, there is a small difference due to OPENNLP-669 |

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 1.5.2 | Tagging Perf 1.5.3 | Tagging Perf 1.6.0 | Comment |
|-------------|---|--------|--------------------|--|--|--|
| Name Finder | CONLL 2002 Dutch Person ned.testa | joern | | Precision: 0.7570754716981132 Recall: 0.4566145092460882 F-Measure: 0.5696539485359361 | Precision: 0.7570754716981132 Recall: 0.4566145092460882 F-Measure: 0.5696539485359361 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Dutch Person ned.testb | joern | | Precision: 0.8479899497487438 Recall: 0.6147540983606558 F-Measure: 0.7127771911298839 | Precision: 0.8479899497487438 Recall: 0.6147540983606558 F-Measure: 0.7127771911298839 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Dutch Organization ned.testa | joern | | Precision: 0.8561872909698997 Recall: 0.37317784256559766 F-Measure: 0.5197969543147207 | Precision: 0.8561872909698997 Recall: 0.37317784256559766 F-Measure: 0.5197969543147207 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Dutch Organization ned.testb | joern | | Precision: 0.783203125 Recall: 0.4546485260770975 F-Measure: 0.5753228120516498 | Precision: 0.783203125 Recall: 0.4546485260770975 F-Measure: 0.5753228120516498 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Dutch Location ned.testa | joern | | Precision: 0.8427947598253275 Recall: 0.40292275574112735 F-Measure: 0.5451977401129944 | Precision: 0.8427947598253275 Recall: 0.40292275574112735 F-Measure: 0.5451977401129944 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Dutch Location ned.testb | joern | | Precision: 0.8827160493827161 Recall: 0.5542635658914729 F-Measure: 0.680952380952381 | Precision: 0.8827160493827161 Recall: 0.5542635658914729 F-Measure: 0.680952380952381 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Dutch Misc ned.testa | joern | | Precision: 0.8354114713216958 Recall: 0.44786096256684493 F-Measure: 0.5831157528285466 | Precision: 0.8354114713216958 Recall: 0.44786096256684493 F-Measure: 0.5831157528285466 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Dutch Misc ned.testb | joern | | Precision: 0.8354114713216958 Recall: 0.44786096256684493 F-Measure: 0.5831157528285466 | Precision: 0.8354114713216958 Recall: 0.44786096256684493 F-Measure: 0.5831157528285466 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |

| | | | | | | |
|-------------|---|--------|--|---|---|--|
| Name Finder | CONLL 2002 Combined ned.testa | joern | | Precision: 0.691407825736184 Recall: 0.6551987767584098 F-Measure: 0.6728164867517175 | Precision: 0.691407825736184 Recall: 0.6551987767584098 F-Measure: 0.6728164867517175 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Dutch Combined ned.testb | joern | | Precision: 0.7128895932382462 Recall: 0.6848515605176351 F-Measure: 0.6985893619774816 | Precision: 0.7128895932382462 Recall: 0.6848515605176351 F-Measure: 0.6985893619774816 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Spanish Person esp.testa | joern | | Precision: 0.9038718291054739 Recall: 0.5540098199672667 F-Measure: 0.686960933536276 | Precision: 0.9038718291054739 Recall: 0.5540098199672667 F-Measure: 0.686960933536276 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Spanish Person esp.testb | joern | | Precision: 0.9063545150501672 Recall: 0.7374149659863946 F-Measure: 0.8132033008252063 | Precision: 0.9063545150501672 Recall: 0.7374149659863946 F-Measure: 0.8132033008252063 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Spanish Organization esp.testa | jkosin | | Precision: 0.8292880258899676 Recall: 0.6029411764705882 F-Measure: 0.6982288828337874 | Precision: 0.8292880258899676 Recall: 0.6029411764705882 F-Measure: 0.6982288828337874 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Spanish Organization esp.testb | joern | | Precision: 0.8031496062992126 Recall: 0.7285714285714285 F-Measure: 0.7640449438202247 | Precision: 0.8031496062992126 Recall: 0.7285714285714285 F-Measure: 0.7640449438202247 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Spanish Location esp.testa | joern | | Precision: 0.7754189944134078 Recall: 0.7052845528455285 F-Measure: 0.7386907929749867 | Precision: 0.7754189944134078 Recall: 0.7052845528455285 F-Measure: 0.7386907929749867 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Spanish Location esp.testb | jkosin | | Precision: 0.8360433604336044 Recall: 0.5691881918819188 F-Measure: 0.6772777167947311 | Precision: 0.8360433604336044 Recall: 0.5691881918819188 F-Measure: 0.6772777167947311 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Spanish Misc esp.testa | jkosin | | Precision: 0.6308411214953271 Recall: 0.30337078651685395 F-Measure: 0.40971168437025796 | Precision: 0.6308411214953271 Recall: 0.30337078651685395 F-Measure: 0.40971168437025796 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Spanish Misc esp.testb | jkosin | | Precision: 0.6763005780346821 Recall: 0.34513274336283184 F-Measure: 0.4570312499999994 | Precision: 0.6763005780346821 Recall: 0.34513274336283184 F-Measure: 0.4570312499999994 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |

| | | | | | | |
|-------------|---|--------|---|---|---|--|
| Name Finder | CONLL 2002 Spanish Combined esp.testa | joern | | Precision: 0.7213977979894687 Recall: 0.6927143185474604 F-Measure: 0.706765154179857 | Precision: 0.7213977979894687 Recall: 0.6927143185474604 F-Measure: 0.706765154179857 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2002 Spanish Combined esp.testb | joern | | Precision: 0.7612574341546304 Recall: 0.7554806070826307 F-Measure: 0.7583580194667795 | Precision: 0.7612574341546304 Recall: 0.7554806070826307 F-Measure: 0.7583580194667795 | Old results couldn't be reproduced. Training 1.53 and 1.6.0-rc2 with the same data file. Test will be automated for RC3. |
| Name Finder | CONLL 2003 English Person eng.testa | jkosin | Precision: 0.9523195876288659 Recall: 0.8023887079261672 F-Measure: 0.8709487330583382 | Precision: 0.9523195876288659 Recall: 0.8023887079261672 F-Measure: 0.8709487330583382 | | |
| Name Finder | CONLL 2003 English Person eng.testb | jkosin | Precision: 0.9391727493917275 Recall: 0.7161410018552876 F-Measure: 0.8126315789473685 | Precision: 0.9391727493917275 Recall: 0.7161410018552876 F-Measure: 0.8126315789473685 | | |
| Name Finder | CONLL 2003 English Organization eng.testa | jkosin | Precision: 0.8768046198267565 Recall: 0.6793437733035048 F-Measure: 0.7655462184873949 | Precision: 0.8768046198267565 Recall: 0.6793437733035048 F-Measure: 0.7655462184873949 | | |
| Name Finder | CONLL 2003 English Organization eng.testb | jkosin | Precision: 0.8435980551053485 Recall: 0.6267308850090307 F-Measure: 0.7191709844559586 | Precision: 0.8435980551053485 Recall: 0.6267308850090307 F-Measure: 0.7191709844559586 | | |
| Name Finder | CONLL 2003 English Location eng.testa | jkosin | Precision: 0.9361421988150099 Recall: 0.7740881872618399 F-Measure: 0.8474374255065554 | Precision: 0.9361421988150099 Recall: 0.7740881872618399 F-Measure: 0.8474374255065554 | | |
| Name Finder | CONLL 2003 English Location eng.testb | jkosin | Precision: 0.9206349206349206 Recall: 0.7302158273381295 F-Measure: 0.8144433299899699 | Precision: 0.9206349206349206 Recall: 0.7302158273381295 F-Measure: 0.8144433299899699 | | |
| Name Finder | CONLL 2003 English Misc eng.testa | jkosin | Precision: 0.9027982326951399 Recall: 0.6648590021691974 F-Measure: 0.7657713928794503 | Precision: 0.9027982326951399 Recall: 0.6648590021691974 F-Measure: 0.7657713928794503 | | |
| Name Finder | CONLL 2003 English Misc eng.testb | jkosin | Precision: 0.8592436974789915 Recall: 0.5826210826210826 F-Measure: 0.6943972835314092 | Precision: 0.8592436974789915 Recall: 0.5826210826210826 F-Measure: 0.6943972835314092 | | |

| | | | | | | |
|-------------|--|--------|---|---|---|--------------------------------|
| Name Finder | CONLL 2003 English Combined eng.testa | jkosin | Precision: 0.861812521618817 Recall: 0.8386065297879501 F-Measure: 0.8500511770726714 | Precision: 0.8640608785887236 Recall: 0.8407943453382699 F-Measure: 0.8522688502217672 | | 1000 iterations OPENNLP-417 |
| Name Finder | CONLL 2003 English Combined eng.testb | jkosin | Precision: 0.8041311831853597 Recall: 0.7857648725212465 F-Measure: 0.7948419450165667 | Precision: 0.8064866823699945 Recall: 0.7880665722379604 F-Measure: 0.7971702337243664 | | 1000 iterations OPENNLP-417 |
| Name Finder | CONLL 2003 German Person deu.testa | joern | Precision: 0.9132653061224489 Recall: 0.2555317630264097 3 F-Measure: 0.3993307306190742 | Precision: 0.9132653061224489 Recall: 0.2555317630264097 3 F-Measure: 0.3993307306190742 | Precision: 0.9132653061224489 Recall: 0.25553176302640973 F-Measure: 0.3993307306190742 | |
| Name Finder | CONLL 2003 German Person deu.testb | joern | Precision: 0.8732106339468303 Recall: 0.3573221757322176 F-Measure: 0.507125890736342 | Precision: 0.8732106339468303 Recall: 0.3573221757322176 F-Measure: 0.507125890736342 | Precision: 0.8732106339468303 Recall: 0.3573221757322176 F-Measure: 0.507125890736342 | |
| Name Finder | CONLL 2003 German Organization deu.testa | joern | Precision: 0.8407224958949097 Recall: 0.4125705076551168 F-Measure: 0.5535135135135135 | Precision: 0.8407224958949097 Recall: 0.4125705076551168 F-Measure: 0.5535135135135135 | Precision: 0.8407224958949097 Recall: 0.4125705076551168 F-Measure: 0.5535135135135135 | |
| Name Finder | CONLL 2003 German Organization deu.testb | joern | Precision: 0.8014705882352942 Recall: 0.4230271668822768 F-Measure: 0.5537679932260795 | Precision: 0.8014705882352942 Recall: 0.4230271668822768 F-Measure: 0.5537679932260795 | Precision: 0.8014705882352942 Recall: 0.4230271668822768 F-Measure: 0.5537679932260795 | |
| Name Finder | CONLL 2003 German Location deu.testa | joern | Precision: 0.7816326530612245 Recall: 0.3243014394580863 7 F-Measure: 0.45840813883901854 | Precision: 0.7816326530612245 Recall: 0.3243014394580863 7 F-Measure: 0.45840813883901854 | Precision: 0.7816326530612245 Recall: 0.32430143945808637 F-Measure: 0.45840813883901854 | |
| Name Finder | CONLL 2003 German Location deu.testb | joern | Precision: 0.8033826638477801 Recall: 0.3671497584541063 F-Measure: 0.5039787798408487 | Precision: 0.8033826638477801 Recall: 0.3671497584541063 F-Measure: 0.5039787798408487 | Precision: 0.8033826638477801 Recall: 0.3671497584541063 F-Measure: 0.5039787798408487 | |
| Name Finder | CONLL 2003 German Misc deu.testa | joern | Precision: 0.7055555555555556 Recall: 0.1257425742574257 4 F-Measure: 0.21344537815126052 | Precision: 0.7055555555555556 Recall: 0.1257425742574257 4 F-Measure: 0.21344537815126052 | Precision: 0.7055555555555556 Recall: 0.12574257425742574 F-Measure: 0.21344537815126052 | |
| Name Finder | CONLL 2003 German Misc deu.testb | joern | Precision: 0.6601307189542484 Recall: 0.1507462686567164 2 F-Measure: 0.2454434993924666 | Precision: 0.6601307189542484 Recall: 0.1507462686567164 2 F-Measure: 0.2454434993924666 | Precision: 0.6601307189542484 Recall: 0.15074626865671642 F-Measure: 0.2454434993924666 | |

| | | | | | | |
|-------------------|---|---------|---|---|---|--|
| Name Finder | CONLL 2003 German Combined deu.testa | joern | Precision: 0.7718859429714857 Recall: 0.319263397475688 F-Measure: 0.4516978922716628 | Precision: 0.7783891945972986 Recall: 0.3219532381543554 5 F-Measure: 0.45550351288056207 | Precision: 0.7783891945972986 Recall: 0.32195323815435545 F-Measure: 0.45550351288056207 | OPENNLP-417 |
| Name Finder | CONLL 2003 German Combined deu.testb | joern | Precision: 0.7467566165023353 Recall: 0.3917778382793357 F-Measure: 0.5139285714285715 | Precision: 0.749351323300467 Recall: 0.3931391233324258 F-Measure: 0.5157142857142857 | Precision: 0.749351323300467 Recall: 0.3931391233324258 F-Measure: 0.5157142857142857 | OPENNLP-417 |
| POS Tagger | CONLL 2006 Danish | joern | Accuracy: 0.9511278195488722 | Accuracy: 0.9512987012987013 | Accuracy: 0.9512987012987013 | Test will be automated for RC3. |
| POS Tagger | CONLL 2006 Dutch | joern | Accuracy: 0.9324977618621307 | Accuracy: 0.93249776 18621307 | Accuracy: 0.9174574753804834 | Retrained 1.5.3 and got same result as for 1.6.0-rc2. Test will be automated for RC3. |
| POS Tagger | CONLL 2006 Portuguese | joern | Accuracy: 0.9659110277825124 | Accuracy: 0.9659110277825124 | Accuracy: 0.9659110277825124 | Test will be automated for RC3. |
| POS Tagger | CONLL 2006 Swedish | joern | Accuracy: 0.9275106082036775 | Accuracy: 0.9275106082036775 | Accuracy: 0.9275106082036775 | Test will be automated for RC3. |
| Chunker | CONLL 2000 | William | Precision: 0.9257575757575758 Recall: 0.9221868187154117 F-Measure: 0.9239687473746113 | Precision: 0.9257575757575758 Recall: 0.9221868187154117 F-Measure: 0.9239687473746113 | Precision: 0.9257575757575758 Recall: 0.9221868187154117 F-Measure: 0.9239687473746113 | Test will be automated for RC3. |
| Sentence Detector | Arvores Deitadas (Floresta Virgem) (10-fold cross-validation) | William | | Precision: 0.9891491491491492 Recall: 0.9894066523820013 F-Measure: 0.9892778840089301 | Precision: 0.9891491491491492 Recall: 0.9894066523820013 F-Measure: 0.9892778840089301 | PERCEPTRON Cutoff 0 Test will be automated for RC3. |
| Tokenizer | Arvores Deitadas (Floresta Virgem) (10-fold cross-validation) | William | | Precision: 0.9995231988260895 Recall: 0.9994542652270997 F-Measure: 0.9994887308380267 | Precision: 0.9995231988260895 Recall: 0.9994542652270997 F-Measure: 0.9994887308380267 | PERCEPTRON Cutoff 0 alphaNumOpt Test will be automated for RC3. |
| Chunker | Arvores Deitadas (10-fold cross-validation) | William | Precision: 0.9404684925220583 Recall: 0.9374181341871635 F-Measure: 0.9389408359191154 | Precision: 0.9562405864042575 Recall: 0.9582419351592844 F-Measure: 0.9572402147035765 | Precision: 0.9562712068584737 Recall: 0.9585035519510575 F-Measure: 0.9573860781121228 | Retrained 1.5.3 and got same result as for 1.6.0-rc2. Test will be automated for RC3. |

Automated Evaluation Testing

The Conll 2002 tests for example can be run with this command:

```
mvn test -DOPENNLP_DATA_DIR=~/.opennlp-test-data/ -Dtest=Conll02*
```

| Unit Tests | Tester | Passed | Comment |
|-------------|---------|--------|---------|
| OntoNotes4* | Jörn | Yes | |
| ConllL2002* | Jörn | Yes | |
| ConllX* | Jörn | Yes | |
| Arvores* | William | Yes | |

| | | | |
|----------|---------|-----|--|
| Conll00* | William | Yes | |
|----------|---------|-----|--|

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 | | | |
| Sentence Detector Trainer | | | |
| Tokenizer ME | | | |
| Tokenizer Trainer | | | |
| Name Finder | | | |
| Name Finder Trainer | | | |
| Chunker | | | |
| Chunker Trainer | | | |
| POS Tagger | | | |
| POS Tagger Trainer | | | |
| Parser | | | |
| createPear.sh | Jörn | | |
| Sample PEAR | Jörn | | |

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 | William | No | JWNL license should be removed. Snowball license is missing. Fixed in trunk! |
| Binary | NOTICE | Jörn | Yes | |
| Binary | README | Jörn | No | Coref mention should be removed. Fixed in trunk! |
| Binary | RELEASE_NOTES.html | William | Yes | |
| Binary | Test signatures: .md5, .sha1, .asc | Jörn | | |
| Binary | JIRA issue list created | William | Yes | |
| Binary | Contains the tools jar and documentation? | Jörn | Yes | RC4 |
| Source | LICENSE | Jörn | Yes | |
| Source | NOTICE | Jörn | Yes | |
| Source | Test signatures: .md5, .sha1, .asc | Jörn | | |
| Source | Can build from source? | Jörn | Yes | RC4 |

Notes about testing

Compatibility tests

The following commands can be used to reproduce the compatibility tests with Leipzig corpus.

```
# Corpus preparation: the following command will create documents from the corpus. Sed is used to remove the
language prefix

sh bin/opennlp DoccatConverter leipzig -data ../eng_news_2010_300K-text/eng_news_2010_300K-sentences.txt -
encoding UTF-8 -lang en | sed -E 's/^en[[:space:]]//g' > ../out-tokenized-documents.test

# Corpus preparation: this forces the detokenization of the documents

sh bin/opennlp SentenceDetectorConverter namefinder -data ../out-tokenized-documents.test -encoding UTF-8 -
detokenizer trunk/opennlp-tools/lang/en/tokenizer/en-detokenizer.xml > ../out-documents.test

# Now the actually tests. Execute it for the previous release and for the current RC. Compare the output using
diff:

time sh bin/opennlp SentenceDetector ../models/en-sent.bin < ../out-documents.test > ../out-sentences_1.5.2.test
time sh bin/opennlp TokenizerME ../models/en-token.bin < ../out-sentences_1.5.2.test > ../out-toks_1.5.2.test
time sh bin/opennlp TokenNameFinder ../models/en-ner-person.bin < ../out-toks_1.5.2.test > ../out-ner_1.5.2.test
time sh bin/opennlp POSTagger ../models/en-pos-maxent.bin < ../out-toks_1.5.2.test > ../out-pos_maxent_1.5.2.
test
time sh bin/opennlp POSTagger ../models/en-pos-perceptron.bin < ../out-toks_1.5.2.test > ../out-pos_pers_1.5.2.
test
time sh bin/opennlp ChunkerME ../models/en-chunker.bin < ../out-pos_pers_1.5.2.test > ../out-chk_1.5.2.test
time sh bin/opennlp Parser ../models/en-parser-chunking.bin < ../out-toks_1.5.2.test > ../out-parse_1.5.2.test
```