

Test 2 – Language Recognition

| | |
|---|---|
| Sources..... | 1 |
| 3 languages - Training | 1 |
| Running..... | 1 |
| English | 1 |
| Norwegian | 2 |
| German..... | 2 |
| Swedish (without training)..... | 3 |
| 4 languages - Swedish training..... | 4 |
| Swedish Runner..... | 4 |
| Bigger Norwegian test | 4 |
| 5 languages - Danish training and testing | 5 |
| Training | 5 |
| Testing..... | 5 |
| Danish test 2 | 6 |
| Re-test Norwegian, English, German and Swedish..... | 7 |

Sources

- Swedish
<http://mobilforum.mkf.se/default.aspx>
- Norwegian
<http://www.mobiltelefon.no/diskuter/viewtopic.php?t=50>
- German
<http://www.dashandyforum.de/fm5-Nokia.html>
- English
<http://www.cellphonehacks.com/viewforum.php?f=1&sid=bdadad4f285f65ce45ab629f94abb0dd>
- Danish
<http://www.mobildebat.dk/forumdisplay.php?f=34>

3 languages - Training

```
[Constructed Vocabulary (1041 words)]
[Training class: English]
[Trained class English (226 entries, 0.3333333333333333)]
[Training class: Norwegian]
[Trained class Norwegian (274 entries, 0.3333333333333333)]
[Training class: German]
[Trained class German (569 entries, 0.3333333333333333)]
[Saved result in "bayesian.data"]
```

Running

English

1. Read 36 words from engtest1.txt

=====

= Table of Probability =

=====
English 1.0

Norwegian 7.369717775831941E-28

German 7.097819588784965E-31

2. Read 29 words from engtest2.txt

=====
= Table of Probability =

=====
English 0.9999999999999991

Norwegian 8.756211253490469E-16

German 1.8020386973936708E-20

3. Read 12 words from engtest3.txt

=====
= Table of Probability =

=====
English 0.9988138487705428

Norwegian 7.724216806225626E-4

German 4.137295488346896E-4

Norwegian

1. Read 44 words from notest1.txt

=====
= Table of Probability =

=====
English 1.6814002119808864E-55

Norwegian 1.0

German 4.840685641581295E-65

2. Read 27 words from notest2.txt

=====
= Table of Probability =

=====
English 1.4124880328104153E-13

Norwegian 0.999999999998552

German 3.62401585927209E-15

3. Read 44 words from notest3.txt

=====
= Table of Probability =

=====
English 2.7324800153504334E-28

Norwegian 1.0

German 3.5405989436236647E-34

German

1. Read 27 words from getest1.txt

=====
= Table of Probability =
=====
English 4.14362086438058E-23
Norwegian 9.863923835913514E-27
German 1.0

2. Read 36 words from getest2.txt

=====
= Table of Probability =
=====
English 2.9390154863639824E-29
Norwegian 4.703305967733239E-30
German 1.0

3. Read 74 words from getest3.txt

=====
= Table of Probability =
=====
English 2.1840457280874096E-61
Norwegian 8.650960678740025E-66
German 1.0

Swedish (without training)

1. Read 39 words from swedish1.txt

=====
= Table of Probability =
=====
English 3.588950643094305E-4
Norwegian 2.3647996993839178E-4
German 0.9994046249657522

2. Read 45 words from swedish2.txt

=====
= Table of Probability =
=====
English 2.0973076763497128E-16
Norwegian 0.9999999999999998
German 1.5567759405753413E-18

3. Read 39 words from swedish3.txt

=====
= Table of Probability =
=====
English 0.5687299228639023
Norwegian 0.3212078772511325
German 0.11006219988496523

4 languages - Swedish training

```
C:\javaproject>java -jar trainer.jar input.txt stopwords.txt
[Constructed Vocabulary (1224words)]
[Training class: English]
[Trained class English (226 entries, 0.25)]
[Training class: Norwegian]
[Trained class Norwegian (274 entries, 0.25)]
[Training class: Swedish]
[Trained class Swedish (211 entries, 0.25)]
[Training class: German]
[Trained class German (569 entries, 0.25)]
[Saved result in "bayesian.data"]
```

Swedish Runner

```
C:\javaproject>java -jar runner.jar swedish1.txt
```

1. Read 39 words from swedish1.txt

```
=====
= Table of Probability =
=====
English 1.969917605747747E-34
Norwegian 2.089096756972563E-35
Swedish 1.0
German 1.87872893229865E-34
```

2. Read 45 words from swedish2.txt

```
=====
= Table of Probability =
=====
English 2.4733034206671982E-24
Norwegian 3.451090050059233E-9
Swedish 0.9999999965489099
German 3.039471716246577E-29
```

3. Read 39 words from swedish3.txt

```
=====
= Table of Probability =
=====
English 1.1511709793105159E-42
Norwegian 7.253435598593943E-44
Swedish 1.0
German 1.4177775930940837E-47
```

Bigger Norwegian test

```
Read 301 words from notest.txt
```

```
=====
```

= Table of Probability =
=====
English 3.5560714810817093E-203
Norwegian 1.0
Swedish 1.1256700373512438E-173
German 7.686231192917445E-231

5 languages - Danish training and testing

Training

[Constructed Vocabulary (1363words)]
[Training class: English]
[Trained class English (226 entries, 0.2)]
[Training class: Norwegian]
[Trained class Norwegian (274 entries, 0.2)]
[Training class: Danish]
[Trained class Danish (189 entries, 0.2)]
[Training class: Swedish]
[Trained class Swedish (211 entries, 0.2)]
[Training class: German]
[Trained class German (569 entries, 0.2)]
[Saved result in "bayesian.data"]

Testing

1. Read 81 words from dantest1.txt

=====
= Table of Probability =
=====
English 8.991965213077333E-23
Norwegian 0.9993278803329411
Danish 6.72119662864171E-4
Swedish 4.194514153924407E-12
German 1.1501825352679012E-16

2. Read 63 words from dantest2.txt

=====
= Table of Probability =
=====
English 3.869005939751965E-15
Norwegian 0.9994281935913211
Danish 5.718048362497284E-4
Swedish 1.5724159663595226E-9
German 9.399448625921188E-15

3. Read 56 words from dantest3.txt

```
=====
= Table of Probability =
=====
English 1.086485923659732E-28
Norwegian 1.3097789019963041E-7
Danish 0.999998690221096
Swedish 2.0430328479055665E-16
German 2.8404063058362033E-27
```

Danish test 2

```
[Constructed Vocabulary (1421 words)]
[Training class: English]
[Trained class English (226 entries, 0.2)]
[Training class: Norwegian]
[Trained class Norwegian (274 entries, 0.2)]
[Training class: Danish]
[Trained class Danish (256 entries, 0.2)]
[Training class: Swedish]
[Trained class Swedish (211 entries, 0.2)]
[Training class: German]
[Trained class German (569 entries, 0.2)]
[Saved result in "bayesian.data"]
```

Read 81 words from dantest1.txt

```
=====
= Table of Probability =
=====
English 1.759958363545157E-33
Norwegian 1.4573538174667556E-11
Danish 0.99999999854265
Swedish 1.059975270976473E-22
German 4.752106703598606E-28
```

Read 63 words from dantest2.txt

```
=====
= Table of Probability =
=====
English 2.4761183897207192E-20
Norwegian 3.884585350666904E-6
Danish 0.9999961154146347
Swedish 1.449891532031296E-14
German 5.454250997344462E-21
```

Read 56 words from dantest3.txt

```
=====
= Table of Probability =
```

=====
English 3.103451368338644E-37
Norwegian 3.3437561752795824E-16
Danish 0.999999999999996
Swedish 6.762427015623431E-25
German 3.756442145949414E-36

Re-test Norwegian, English, German and Swedish

1. Read 44 words from notest1.txt

=====
= Table of Probability =
=====
English 5.745695064286175E-56
Norwegian 1.0
Danish 2.2012177415333575E-25
Swedish 1.9617597350791415E-57
German 3.591778299242662E-66

2. Read 27 words from notest2.txt

=====
= Table of Probability =
=====
English 1.2069237605534526E-13
Norwegian 0.9989856571824042
Danish 0.0010143426094365341
Swedish 2.0803710928992605E-10
German 1.456335687303482E-15

3. Read 44 words from notest3.txt

=====
= Table of Probability =
=====
English 1.6820740262129093E-28
Norwegian 1.0
Danish 9.681599170501704E-17
Swedish 6.475159654819547E-25
German 9.047559247411E-35

4. Read 36 words from engtest1.txt

=====
= Table of Probability =
=====
English 1.0
Norwegian 3.908468248475566E-28
Danish 9.371625794009727E-29
Swedish 6.879948304927391E-28
German 3.9280632799923035E-31

5. Read 29 words from engtest2.txt

=====
= Table of Probability =
=====

English 0.999999999968541

Norwegian 6.1849241700157E-16

Danish 8.29115591511084E-19

Swedish 3.1453935454898702E-12

German 1.17238521113408E-20

6. Read 12 words from engtest3.txt

=====
= Table of Probability =
=====

English 0.9972220774835818

Norwegian 7.17870578591789E-4

Danish 7.065272511648306E-4

Swedish 9.690141556223538E-4

German 3.8451053103933665E-4

7. Read 27 words from getest1.txt

=====
= Table of Probability =
=====

English 1.4474645419098348E-23

Norwegian 3.2074659616540764E-27

Danish 2.7791175939695552E-27

Swedish 4.772190355288342E-26

German 1.0

8. Read 36 words from getest2.txt

=====
= Table of Probability =
=====

English 8.435376267482436E-30

Norwegian 1.3499131162460097E-30

Danish 7.775267484991104E-28

Swedish 5.70802116191144E-26

German 1.0

9. Read 74 words from getest3.txt

=====
= Table of Probability =
=====

English 3.8255215415798534E-62

Norwegian 1.1743505845143648E-66

Danish 5.587280131318678E-64

Swedish 9.085631809822178E-61

German 1.0

10. Read 39 words from swedish1.txt

=====
= Table of Probability =
=====

English 1.2725098646619434E-34

Norwegian 1.129030896015264E-35

Danish 5.635382906137376E-27

Swedish 1.0

German 5.872773926089439E-35

11. Read 45 words from swedish2.txt

=====
= Table of Probability =
=====

English 1.7498383284440945E-24

Norwegian 2.470980434864447E-9

Danish 9.302019610150826E-17

Swedish 0.999999975290195

German 1.0076045386324478E-29

12. Read 39 words from swedish3.txt

=====
= Table of Probability =
=====

English 7.880319049922869E-43

Norwegian 4.989555677040734E-44

Danish 1.325761140639008E-44

Swedish 1.0

German 9.948478622422859E-48