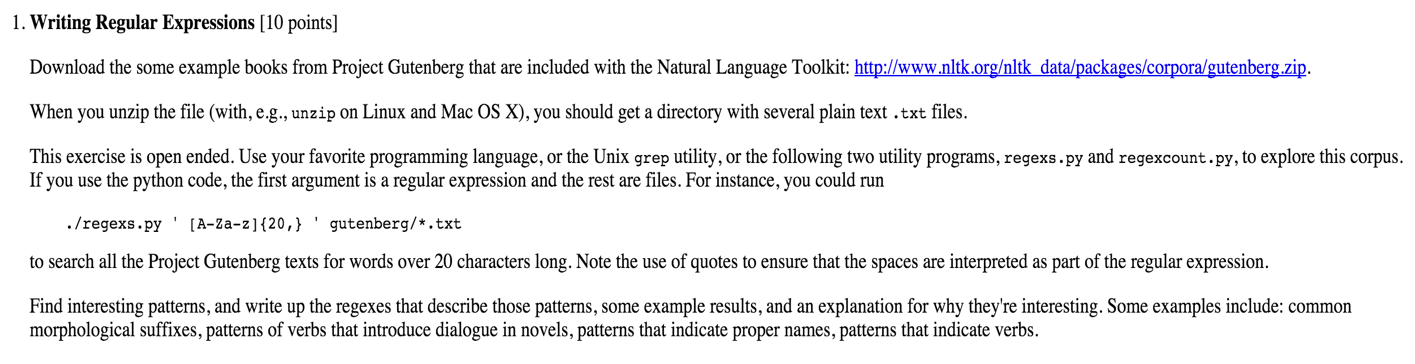
CS 6120 Assignment 1

NUID: 001669685 DHEERAJ JOSHI

Q.1)



Solution:

Unzip the REGEX.zip file from the unzipped HW1 folder.

The project can be directly imported to Pycharm or executed on the python console.

The regexs.py file runs the regex patterns on the Gutenberg corpus and returns the matching result set in separate .csv files.

The regexcount.py returns the count of the corresponding patterns on the console.

Regex Patterns:

1. “[A-Z][a-z]+[ ]\*”: The pattern detects Pronouns by find all word that start with an uppercase letter. Although, pronouns with acronyms and special characters are not detected.

Sample results: Hashubah  
 Omnipotent  
 Doeg  
 Justification  
 Debts  
 Lambent

1. “([\w\-\.]+@(\w[\w\-]+\.)+[\w\-]+)” : The pattern detects the Email ID’s present in the Gutenberg corpus.

Sample results: hart@vmd.cso.uiuc.edu  
 crowderb@blankslate.net  
 hart@pobox.com  
 72600.2026@compuserve.com

1. “^(.\*?)(ing|ly|ed|ious|ies|ive|es|s|ment)$” : Some Common morphological suffixes used for stemming are detected by this pattern.

Sample results: unimaginative  
 hordes  
 troubledly  
 Hasting

1. “(\d+)” : All words representing digits are detected by this pattern.

Sample results: 1788  
 1789  
 212  
 217  
 165  
 1818

1. “\$[0-9.,]+”: All words that depict price or cost are detected by this pattern.

Sample results: $7,000,000.  
 $20,000,000  
 $2

Q.2)

a.) Unvowel.txt

0 0 <space> <space> 1

0 0 ! ! 1

0 0 " " 1

0 0 # # 1

0 0 $ $ 1

0 0 % % 1

0 0 & & 1

0 0 ' ' 1

0 0 ( ( 1

0 0 ) ) 1

0 0 \* \* 1

0 0 + + 1

0 0 , , 1

0 0 - - 1

0 0 . . 1

0 0 / / 1

0 0 0 0 1

0 0 1 1 1

0 0 2 2 1

0 0 3 3 1

0 0 4 4 1

0 0 5 5 1

0 0 6 6 1

0 0 7 7 1

0 0 8 8 1

0 0 9 9 1

0 0 : : 1

0 0 ; ; 1

0 0 < < 1

0 0 = = 1

0 0 > > 1

0 0 ? ? 1

0 0 @ @ 1

0 0 A <epsilon> 1

0 0 B B 1

0 0 C C 1

0 0 D D 1

0 0 E <epsilon> 1

0 0 F F 1

0 0 G G 1

0 0 H H 1

0 0 I <epsilon> 1

0 0 J J 1

0 0 K K 1

0 0 L L 1

0 0 M M 1

0 0 N N 1

0 0 O <epsilon> 1

0 0 P P 1

0 0 Q Q 1

0 0 R R 1

0 0 S S 1

0 0 T T 1

0 0 U <epsilon> 1

0 0 V V 1

0 0 W W 1

0 0 X X 1

0 0 Y Y 1

0 0 Z Z 1

0 0 [ [ 1

0 0 \ \ 1

0 0 ] ] 1

0 0 ^ ^ 1

0 0 \_ \_ 1

0 0 ` ` 1

0 0 a <epsilon> 1

0 0 b b 1

0 0 c c 1

0 0 d d 1

0 0 e <epsilon> 1

0 0 f f 1

0 0 g g 1

0 0 h h 1

0 0 i <epsilon> 1

0 0 j j 1

0 0 k k 1

0 0 l l 1

0 0 m m 1

0 0 n n 1

0 0 o <epsilon> 1

0 0 p p 1

0 0 q q 1

0 0 r r 1

0 0 s s 1

0 0 t t 1

0 0 u <epsilon> 1

0 0 v v 1

0 0 w w 1

0 0 x x 1

0 0 y y 1

0 0 z z 1

0 0 { { 1

0 0 | | 1

0 0 } } 1

0 0 ~ ~ 1

1. Compile Unvowel.txt

**fstcompile --isymbols=ascii.syms --osymbols=ascii.syms < unvowel.txt > unvowel.fst**

2. Compose it with Declaration.fst

**fstcompose declaration.fst unvowel.fst | fstproject --project\_output | fstrmepsilon | fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 1

1 2 h h 1

2 3 n n 2

3 4 <space> <space> 1

4 5 n n 2

5 6 <space> <space> 1

6 7 t t 1

7 8 h h 1

8 9 <space> <space> 2

9 10 c c 1

10 11 r r 3

11 12 s s 1

12 13 <space> <space> 2

13 14 f f 2

14 15 <space> <space> 1

15 16 h h 1

16 17 m m 2

17 18 n n 2

18 19 <space> <space> 1

19 20 v v 2

20 21 n n 2

21 22 t t 1

22 23 s s 1

23 24 <space> <space> 1

24 25 t t 2

25 26 <space> <space> 1

26 27 b b 1

27 28 c c 2

28 29 m m 2

29 30 s s 2

30 31 <space> <space> 1

31 32 n n 1

32 33 c c 2

33 34 s s 2

34 35 s s 1

35 36 r r 2

36 37 y y 1

37 38 <space> <space> 1

38 39 f f 1

39 40 r r 2

40 41 <space> <space> 1

41 42 n n 2

42 43 <space> <space> 2

43 44 p p 1

44 45 p p 3

45 46 l l 1

46 47 <space> <space> 2

47 48 t t 1

48 49 <space> <space> 2

49 50 d d 1

50 51 s s 2

51 52 s s 1

52 53 l l 2

53 54 v v 1

54 55 <space> <space> 2

55 56 t t 1

56 57 h h 1

57 58 <space> <space> 2

58 59 p p 1

59 60 l l 2

60 61 t t 2

61 62 c c 2

62 63 l l 2

63 64 <space> <space> 1

64 65 b b 1

65 66 n n 2

66 67 d d 1

67 68 s s 1

68 69 <space> <space> 1

3. Invert the Transducer

**fstinvert unvowel.fst invert-vowel.fst**

4. Compose Inverted Transducer with the given c1, c2, c3,c5, c7 n-gram language models

**fstcompose invert-vowel.fst c1.fst revowel1.fst**

**fstcompose invert-vowel.fst c2.fst revowel2.fst**

**fstcompose invert-vowel.fst c3.fst revowel3.fst**

**fstcompose invert-vowel.fst c5.fst revowel5.fst**

**fstcompose invert-vowel.fst c7.fst revowel7.fst**

5. Compose Declaration.fst with each of the Revowel Transducers

**fstcompose declaration.fst unvowel.fst | fstproject --project\_output | fstrmepsilon | fstcompose – revowel1.fst | fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort | fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 5.96621704

1 2 h h 4.98792267

2 3 n n 5.82987309

3 4 <space> <space> 3.70899105

4 5 n n 5.82987309

5 6 <space> <space> 3.70899105

6 7 t t 4.5414238

7 8 h h 4.98792267

8 9 <space> <space> 4.70899105

9 10 c c 5.96735191

10 11 r r 7.03404522

11 12 s s 5.00382519

12 13 <space> <space> 4.70899105

13 14 f f 6.98733997

14 15 <space> <space> 3.70899105

15 16 h h 4.98792267

16 17 m m 6.86242962

17 18 n n 5.82987309

18 19 <space> <space> 3.70899105

19 20 v v 7.92842627

20 21 n n 5.82987309

21 22 t t 4.5414238

22 23 s s 5.00382519

23 24 <space> <space> 3.70899105

24 25 t t 5.5414238

25 26 <space> <space> 3.70899105

26 27 b b 6.39844704

27 28 c c 6.96735191

28 29 m m 6.86242962

29 30 s s 6.00382519

30 31 <space> <space> 3.70899105

31 32 n n 4.82987309

32 33 c c 6.96735191

33 34 s s 6.00382519

34 35 s s 5.00382519

35 36 r r 6.03404522

36 37 y y 6.22872734

37 38 <space> <space> 3.70899105

38 39 f f 5.98733997

39 40 r r 6.03404522

40 41 <space> <space> 3.70899105

41 42 n n 5.82987309

42 43 <space> <space> 4.70899105

43 44 p p 6.24422455

44 45 p p 8.24422455

45 46 l l 5.46552086

46 47 <space> <space> 4.70899105

47 48 t t 4.5414238

48 49 <space> <space> 4.70899105

49 50 d d 5.23390865

50 51 s s 6.00382519

51 52 s s 5.00382519

52 53 l l 6.46552086

53 54 v v 6.92842627

54 55 <space> <space> 4.70899105

55 56 t t 4.5414238

56 57 h h 4.98792267

57 58 <space> <space> 4.70899105

58 59 p p 6.24422455

59 60 l l 6.46552086

60 61 t t 5.5414238

61 62 c c 6.96735191

62 63 l l 6.46552086

63 64 <space> <space> 3.70899105

64 65 b b 6.39844704

65 66 n n 5.82987309

66 67 d d 5.23390865

67 68 s s 5.00382519

68 69 <space> <space> 3.70899105

**fstcompose declaration.fst unvowel.fst | fstproject --project\_output | fstrmepsilon | fstcompose - revowel2.fst |fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort |fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 4.60410023

1 2 h h 4.27439594

2 3 e e 1.68769312

3 4 n n 5.52267504

4 5 <space> <space> 3.41942358

5 6 n n 6.95673466

6 7 <space> <space> 3.41942358

7 8 t t 3.75658226

8 9 h h 3.00378275

9 10 <space> <space> 5.49277496

10 11 c c 5.2437067

11 12 r r 6.54969501

12 13 s s 5.00708199

13 14 <space> <space> 3.97956038

14 15 f f 6.23334122

15 16 <space> <space> 2.85894156

16 17 h h 4.86565876

17 18 e e 1.68769312

18 19 m m 6.89795065

19 20 a a 2.77111197

20 21 n n 4.55948925

21 22 <space> <space> 3.41942358

22 23 v v 7.95190716

23 24 e e 1.38632178

24 25 n n 5.52267504

25 26 t t 4.4496727

26 27 s s 5.74970436

27 28 <space> <space> 2.97956038

28 29 t t 4.75658226

29 30 <space> <space> 3.45042324

30 31 b b 5.08396196

31 32 u u 3.0589993

32 33 c c 6.43820763

33 34 o o 2.84483671

34 35 m m 5.93678665

35 36 s s 6.9670105

36 37 <space> <space> 2.97956038

37 38 n n 5.95673466

38 39 c c 6.56931734

39 40 a a 3.10105991

40 41 s s 5.35608196

41 42 s s 5.06071281

42 43 e e 3.14921284

43 44 r r 5.0707159

44 45 y y 5.41473103

45 46 <space> <space> 2.25662518

46 47 f f 5.23334122

47 48 r r 5.57429409

48 49 <space> <space> 3.63552785

49 50 n n 6.95673466

50 51 <space> <space> 4.41942358

51 52 p p 5.55274916

52 53 p p 6.71315861

53 54 l l 4.24203587

54 55 <space> <space> 5.20587301

55 56 t t 3.75658226

56 57 <space> <space> 4.45042324

57 58 d d 5.51577187

58 59 s s 6.46498394

59 60 s s 5.06071281

60 61 l l 7.36968231

61 62 v v 6.91102743

62 63 e e 1.38632178

63 64 <space> <space> 4.02307367

64 65 t t 3.75658226

65 66 h h 3.00378275

66 67 <space> <space> 5.49277496

67 68 p p 5.55274916

68 69 l l 5.24203587

69 70 t t 6.97913122

70 71 c c 8.08190918

71 72 l l 5.8143816

72 73 <space> <space> 4.20587301

73 74 b b 5.08396196

74 75 e e 2.31731367

75 76 n n 5.52267504

76 77 d d 3.60656905

77 78 s s 5.46498394

78 79 <space> <space> 2.97956038

**fstcompose declaration.fst unvowel.fst | fstproject --project\_output | fstrmepsilon | fstcompose - revowel3.fst | fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort |fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 4.60410023

1 2 h h 2.66454315

2 3 i i 2.00102377

3 4 n n 4.16762257

4 5 <space> <space> 3.50232553

5 6 a a 2.81300926

6 7 n n 3.95423698

7 8 <space> <space> 4.00359726

8 9 t t 3.34272814

9 10 h h 2.28522635

10 11 e e 1.35988283

11 12 <space> <space> 3.45336318

12 13 c c 4.87596464

13 14 r r 6.10458231

14 15 e e 2.34316754

15 16 s s 4.35625696

16 17 <space> <space> 3.78471208

17 18 o o 3.28944659

18 19 f f 3.57291579

19 20 <space> <space> 2.04530096

20 21 h h 5.0115099

21 22 a a 2.05648804

22 23 m m 6.21440411

23 24 e e 1.82324862

24 25 n n 4.63849545

25 26 <space> <space> 2.91914487

26 27 v v 8.11191559

27 28 e e 2.07010531

28 29 n n 5.15902996

29 30 t t 3.26032686

30 31 s s 4.88752651

31 32 <space> <space> 2.14498758

32 33 t t 4.92173195

33 34 o o 2.88887739

34 35 <space> <space> 2.45522833

35 36 b b 4.89150953

36 37 e e 2.08525467

37 38 c c 5.96778488

38 39 a a 3.0632267

39 40 m m 4.40555239

40 41 s s 6.66972923

41 42 <space> <space> 2.51859236

42 43 a a 2.7957592

43 44 n n 2.95423698

44 45 c c 6.26208782

45 46 e e 1.45130849

46 47 s s 5.04981804

47 48 s s 3.71643329

48 49 e e 2.63232446

49 50 r r 6.06487942

50 51 y y 4.80678463

51 52 <space> <space> 2.27946734

52 53 f f 5.17545509

53 54 o o 2.2566762

54 55 r r 3.36134315

55 56 <space> <space> 3.02043962

56 57 a a 2.83805418

57 58 n n 3.95423698

58 59 <space> <space> 5.00359726

59 60 p p 6.02298069

60 61 e e 2.65750647

61 62 o o 2.81087565

62 63 p p 4.25629711

63 64 l l 3.09632874

64 65 e e 1.71689034

65 66 <space> <space> 3.90927219

66 67 t t 4.19441509

67 68 o o 2.88887739

68 69 <space> <space> 3.45522833

69 70 d d 5.1143589

70 71 e e 2.52583623

71 72 s s 5.18131256

72 73 s s 3.71643329

73 74 e e 2.63232446

74 75 l l 5.38525581

75 76 v v 5.54194832

76 77 e e 1.42412555

77 78 <space> <space> 4.14656782

78 79 t t 4.19441509

79 80 h h 2.28522635

80 81 e e 1.35988283

81 82 <space> <space> 3.45336318

82 83 p p 5.1651535

83 84 l l 5.40753746

84 85 a a 2.04834318

85 86 t t 5.21706676

86 87 i i 3.22607231

87 88 c c 5.86377907

88 89 a a 2.97710586

89 90 l l 5.00540686

90 91 <space> <space> 3.56489825

91 92 b b 4.99757814

92 93 a a 3.55952811

93 94 n n 4.64118338

94 95 d d 2.54704547

95 96 s s 5.84989262

96 97 <space> <space> 2.11138487

97 98 i i 3.44708538

**fstcompose declaration.fst unvowel.fst | fstproject --project\_output | fstrmepsilon | fstcompose - revowel5.fst | fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort |fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 4.60410023

1 2 h h 2.66454315

2 3 e e 2.35879779

3 4 n n 3.55122757

4 5 <space> <space> 2.04214239

5 6 i i 2.37982273

6 7 n n 4.07805824

7 8 <space> <space> 2.48634529

8 9 t t 2.80491948

9 10 h h 2.06189346

10 11 e e 1.15648305

11 12 <space> <space> 3.2382586

12 13 c c 4.45852709

13 14 o o 2.14541817

14 15 u u 2.4502182

15 16 r r 6.11219978

16 17 s s 2.72508955

17 18 <space> <space> 3.65889549

18 19 o o 2.72567987

19 20 f f 3.38204408

20 21 <space> <space> 2.00006342

21 22 h h 5.00897503

22 23 u u 2.68937397

23 24 m m 3.6347115

24 25 a a 1.20132816

25 26 n n 3.00035

26 27 <space> <space> 2.20361757

27 28 v v 7.69812775

28 29 e e 2.14472723

29 30 n n 4.51014233

30 31 t t 2.74299836

31 32 s s 4.20740795

32 33 <space> <space> 2.00066233

33 34 t t 5.05186701

34 35 o o 2.60451078

35 36 <space> <space> 2.26126122

36 37 b b 4.61446762

37 38 e e 1.26767516

38 39 c c 6.14478922

39 40 o o 2.31792307

40 41 m m 3.04758978

41 42 e e 1.22527313

42 43 s s 5.6012125

43 44 <space> <space> 2.01255655

44 45 n n 6.1854682

45 46 e e 1.64390683

46 47 c c 5.44074965

47 48 e e 1.43731618

48 49 s s 3.06878734

49 50 s s 2.28544641

50 51 a a 2.51453686

51 52 r r 3.78899503

52 53 y y 3.58357859

53 54 <space> <space> 2.01008129

54 55 f f 4.42279482

55 56 o o 2.16488862

56 57 r r 3.37866712

57 58 <space> <space> 2.32046366

58 59 i i 3.6746459

59 60 n n 4.61964512

60 61 <space> <space> 3.54378057

61 62 p p 6.22939253

62 63 e e 3.54055023

63 64 o o 1.84008741

64 65 p p 4.00000906

65 66 l l 2.00009012

66 67 e e 1.00002003

67 68 <space> <space> 3.02203989

68 69 t t 4.43917561

69 70 o o 2.04503608

70 71 <space> <space> 3.42009258

71 72 d d 5.08002949

72 73 i i 2.13500357

73 74 s s 3.7211287

74 75 s s 5.60415268

75 76 o o 3.0116961

76 77 l l 3.5579617

77 78 v v 2.60895443

78 79 e e 1.00270414

79 80 <space> <space> 5.02176666

80 81 t t 3.72435093

81 82 h h 2.34322095

82 83 e e 1.55643368

83 84 <space> <space> 3.2382586

84 85 p p 4.67258644

85 86 o o 2.86510992

86 87 l l 5.25078917

87 88 i i 1.50907183

88 89 t t 5.28289318

89 90 i i 2.82918549

90 91 c c 4.88956642

91 92 a a 2.08810735

92 93 l l 3.21859312

93 94 <space> <space> 2.33520889

94 95 b b 4.48344707

95 96 o o 2.41894817

96 97 u u 5.11327791

97 98 n n 4.76736832

98 99 d d 2.09768295

99 100 s s 4.71538639

100 101 <space> <space> 2.00610209

101 102 i i 3.92292905

**fstcompose declaration.fst unvowel.fst | fstproject --project\_output | fstrmepsilon | fstcompose - revowel7.fst | fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort |fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 4.60410023

1 2 h h 2.66454315

2 3 e e 2.35879779

3 4 n n 3.55122757

4 5 <space> <space> 2.00591278

5 6 i i 1.75286901

6 7 n n 8.09879684

7 8 <space> <space> 2.17637062

8 9 t t 3.1464448

9 10 h h 2.06371593

10 11 e e 1.08522701

11 12 <space> <space> 3.07994556

12 13 c c 4.41605091

13 14 o o 2.13718605

14 15 u u 3.07013893

15 16 r r 6.06479836

16 17 s s 2.82562995

17 18 e e 1.00237775

18 19 <space> <space> 3.11576366

19 20 o o 3.01724505

20 21 f f 3.03536487

21 22 <space> <space> 2.0073452

22 23 h h 4.32929659

23 24 u u 2.39522171

24 25 m m 3.56074691

25 26 a a 1.00107956

26 27 n n 3.00000048

27 28 <space> <space> 2.21061182

28 29 v v 6.66765261

29 30 e e 2.58680749

30 31 n n 3.21664023

31 32 t t 2.22924089

32 33 s s 6.91545773

33 34 <space> <space> 2.00013256

34 35 t t 4.08710575

35 36 o o 2.17102361

36 37 <space> <space> 2.44294596

37 38 b b 5.20268393

38 39 e e 1.22946274

39 40 c c 7.19264984

40 41 o o 1.28297567

41 42 m m 3.00096869

42 43 e e 1.14419329

43 44 s s 10.8793268

44 45 <space> <space> 2.00178432

45 46 n n 8.48805428

46 47 e e 1.27099776

47 48 c c 3.57436132

48 49 e e 2.65568495

49 50 s s 3.00302601

50 51 s s 2.00205421

51 52 a a 1.70883012

52 53 r r 3.07062626

53 54 y y 2.56603074

54 55 <space> <space> 2.00062704

55 56 f f 3.30133438

56 57 o o 1.33036757

57 58 r r 3.34696341

58 59 <space> <space> 2.262146

59 60 o o 7.117383

60 61 n n 3.05086637

61 62 e e 1.11323619

62 63 <space> <space> 3.02099943

63 64 p p 4.19642735

64 65 e e 2.88093567

65 66 o o 1.65696752

66 67 p p 4

67 68 l l 2

68 69 e e 1

69 70 <space> <space> 3.01904869

70 71 t t 4.5633707

71 72 o o 2.13061786

72 73 <space> <space> 3.21024466

73 74 d d 4.20483541

74 75 i i 2.11217284

75 76 s s 3.7410965

76 77 s s 8.16910076

77 78 o o 2.74205208

78 79 l l 3.11306381

79 80 v v 2.12106109

80 81 e e 1.00067532

81 82 <space> <space> 5.07286644

82 83 t t 2.86851978

83 84 h h 3.72951531

84 85 e e 1.40056932

85 86 <space> <space> 3.37881565

86 87 p p 4.71044064

87 88 o o 2.86137486

88 89 l l 4.99965429

89 90 i i 1.18670535

90 91 t t 5.03669834

91 92 i i 1.23555017

92 93 c c 3.41130567

93 94 a a 1.18126607

94 95 l l 3.0503335

95 96 <space> <space> 2.21470952

96 97 b b 3.76300621

97 98 o o 2.39221263

98 99 u u 6.9050374

99 100 n n 4.76736832

100 101 d d 2.04764915

101 102 s s 6.10168076

102 103 <space> <space> 2.0003581

103 104 i i 2.77392077

b.)

Squash.txt

1. Compile Squash.txt

**fstcompile --isymbols=ascii.syms --osymbols=ascii.syms < squash.txt > squash.fst**

2. Compose Squash.fst with Declaration.fst

**fstcompose declaration.fst squash.fst | fstproject --project\_output | fstrmepsilon | fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 1

1 2 h h 1

2 3 n n 2

3 4 <space> <space> 1

4 5 i i 1

5 6 n n 1

6 7 <space> <space> 1

7 8 t t 1

8 9 h h 1

9 10 e e 1

10 11 <space> <space> 1

11 12 c c 1

12 13 r r 3

13 14 s s 1

14 15 e e 1

15 16 <space> <space> 1

16 17 o o 1

17 18 f f 1

18 19 <space> <space> 1

19 20 h h 1

20 21 m m 2

21 22 n n 2

22 23 <space> <space> 1

23 24 e e 1

24 25 v v 1

25 26 n n 2

26 27 t t 1

27 28 s s 1

28 29 <space> <space> 1

29 30 i i 1

30 31 t t 1

31 32 <space> <space> 1

32 33 b b 1

33 34 c c 2

34 35 m m 2

35 36 s s 2

36 37 <space> <space> 1

37 38 n n 1

38 39 c c 2

39 40 s s 2

40 41 s s 1

41 42 r r 2

42 43 y y 1

43 44 <space> <space> 1

44 45 f f 1

45 46 r r 2

46 47 <space> <space> 1

47 48 o o 1

48 49 n n 1

49 50 e e 1

50 51 <space> <space> 1

51 52 p p 1

52 53 p p 3

53 54 l l 1

54 55 e e 1

55 56 <space> <space> 1

56 57 t t 1

57 58 o o 1

58 59 <space> <space> 1

59 60 d d 1

60 61 s s 2

61 62 s s 1

62 63 l l 2

63 64 v v 1

64 65 e e 1

65 66 <space> <space> 1

66 67 t t 1

67 68 h h 1

68 69 e e 1

69 70 <space> <space> 1

70 71 p p 1

71 72 l l 2

72 73 t t 2

73 74 c c 2

74 75 l l 2

75 76 <space> <space> 1

76 77 b b 1

77 78 n n 2

78 79 d d 1

79 80 s s 1

80 81 <space> <space> 1

3. Invert the Squash Transducer

**fstinvert squash.fst invert-squash.fst**

4. Compose Inverted Transducer with the given c1, c2, c3,c5, c7 n-gram language models

**fstcompose invert-squash.fst c1.fst resquash1.fst**

**fstcompose invert-squash.fst c2.fst resquash2.fst**

**fstcompose invert-squash.fst c3.fst resquash3.fst**

**fstcompose invert-squash.fst c5.fst resquash5.fst**

**fstcompose invert-squash.fst c7.fst resquash7.fst**

5. Compose Declaration.fst with each of the Resquash Transducers

**fstcompose declaration.fst squash.fst | fstproject --project\_output | fstrmepsilon | fstcompose - resquash1.fst |fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort |fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 5.96621704

1 2 h h 4.98792267

2 3 n n 5.82987309

3 4 <space> <space> 3.70899105

4 5 i i 4.89536858

5 6 n n 4.82987309

6 7 <space> <space> 3.70899105

7 8 t t 4.5414238

8 9 h h 4.98792267

9 10 e e 4.2783432

10 11 <space> <space> 3.70899105

11 12 c c 5.96735191

12 13 r r 7.03404522

13 14 s s 5.00382519

14 15 e e 4.2783432

15 16 <space> <space> 3.70899105

16 17 o o 4.86243343

17 18 f f 5.98733997

18 19 <space> <space> 3.70899105

19 20 h h 4.98792267

20 21 m m 6.86242962

21 22 n n 5.82987309

22 23 <space> <space> 3.70899105

23 24 e e 4.2783432

24 25 v v 6.92842627

25 26 n n 5.82987309

26 27 t t 4.5414238

27 28 s s 5.00382519

28 29 <space> <space> 3.70899105

29 30 i i 4.89536858

30 31 t t 4.5414238

31 32 <space> <space> 3.70899105

32 33 b b 6.39844704

33 34 c c 6.96735191

34 35 m m 6.86242962

35 36 s s 6.00382519

36 37 <space> <space> 3.70899105

37 38 n n 4.82987309

38 39 c c 6.96735191

39 40 s s 6.00382519

40 41 s s 5.00382519

41 42 r r 6.03404522

42 43 y y 6.22872734

43 44 <space> <space> 3.70899105

44 45 f f 5.98733997

45 46 r r 6.03404522

46 47 <space> <space> 3.70899105

47 48 o o 4.86243343

48 49 n n 4.82987309

49 50 e e 4.2783432

50 51 <space> <space> 3.70899105

51 52 p p 6.24422455

52 53 p p 8.24422455

53 54 l l 5.46552086

54 55 e e 4.2783432

55 56 <space> <space> 3.70899105

56 57 t t 4.5414238

57 58 o o 4.86243343

58 59 <space> <space> 3.70899105

59 60 d d 5.23390865

60 61 s s 6.00382519

61 62 s s 5.00382519

62 63 l l 6.46552086

63 64 v v 6.92842627

64 65 e e 4.2783432

65 66 <space> <space> 3.70899105

66 67 t t 4.5414238

67 68 h h 4.98792267

68 69 e e 4.2783432

69 70 <space> <space> 3.70899105

70 71 p p 6.24422455

71 72 l l 6.46552086

72 73 t t 5.5414238

73 74 c c 6.96735191

74 75 l l 6.46552086

75 76 <space> <space> 3.70899105

76 77 b b 6.39844704

77 78 n n 5.82987309

78 79 d d 5.23390865

79 80 s s 5.00382519

80 81 <space> <space> 3.70899105

**fstcompose declaration.fst squash.fst | fstproject --project\_output | fstrmepsilon | fstcompose – resquash2.fst |fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort |fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 4.60410023

1 2 h h 4.27439594

2 3 e e 1.68769312

3 4 n n 5.52267504

4 5 <space> <space> 3.41942358

5 6 i i 4.70741653

6 7 n n 3.15188503

7 8 <space> <space> 3.41942358

8 9 t t 3.75658226

9 10 h h 3.00378275

10 11 e e 2.68769312

11 12 <space> <space> 3.02307367

12 13 c c 5.2437067

13 14 r r 6.54969501

14 15 s s 5.00708199

15 16 e e 4.14921284

16 17 <space> <space> 3.02307367

17 18 o o 4.65158081

18 19 f f 4.04879189

19 20 <space> <space> 2.85894156

20 21 h h 4.86565876

21 22 e e 1.68769312

22 23 m m 6.89795065

23 24 a a 2.77111197

24 25 n n 4.55948925

25 26 <space> <space> 3.41942358

26 27 e e 5.96102095

27 28 v v 6.37813616

28 29 e e 1.38632178

29 30 n n 5.52267504

30 31 t t 4.4496727

31 32 s s 5.74970436

32 33 <space> <space> 2.97956038

33 34 i i 4.70741653

34 35 t t 4.01930618

35 36 <space> <space> 3.45042324

36 37 b b 5.08396196

37 38 u u 3.0589993

38 39 c c 6.43820763

39 40 o o 2.84483671

40 41 m m 5.93678665

41 42 s s 6.9670105

42 43 <space> <space> 2.97956038

43 44 n n 5.95673466

44 45 c c 6.56931734

45 46 a a 3.10105991

46 47 s s 5.35608196

47 48 s s 5.06071281

48 49 e e 3.14921284

49 50 r r 5.0707159

50 51 y y 5.41473103

51 52 <space> <space> 2.25662518

52 53 f f 5.23334122

53 54 r r 5.57429409

54 55 <space> <space> 3.63552785

55 56 o o 4.65158081

56 57 n n 3.89795971

57 58 e e 4.47610664

58 59 <space> <space> 3.02307367

59 60 p p 5.55274916

60 61 p p 6.71315861

61 62 l l 4.24203587

62 63 e e 3.70818067

63 64 <space> <space> 3.02307367

64 65 t t 3.75658226

65 66 o o 4.4768362

66 67 <space> <space> 4.18763447

67 68 d d 5.51577187

68 69 s s 6.46498394

69 70 s s 5.06071281

70 71 l l 7.36968231

71 72 v v 6.91102743

72 73 e e 2.38632178

73 74 <space> <space> 3.02307367

74 75 t t 3.75658226

75 76 h h 3.00378275

76 77 e e 2.68769312

77 78 <space> <space> 3.02307367

78 79 p p 5.55274916

79 80 l l 5.24203587

80 81 t t 6.97913122

81 82 c c 8.08190918

82 83 l l 5.8143816

83 84 <space> <space> 4.20587301

84 85 b b 5.08396196

85 86 e e 2.31731367

86 87 n n 5.52267504

87 88 d d 3.60656905

88 89 s s 5.46498394

89 90 <space> <space> 2.97956038

**fstcompose declaration.fst squash.fst | fstproject --project\_output | fstrmepsilon | fstcompose – resquash3.fst |fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort |fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 4.60410023

1 2 h h 2.66454315

2 3 i i 2.00102377

3 4 n n 4.16762257

4 5 <space> <space> 3.50232553

5 6 i i 4.57923079

6 7 n n 2.93230724

7 8 <space> <space> 3.50232553

8 9 t t 3.34272814

9 10 h h 2.28522635

10 11 e e 2.35988283

11 12 <space> <space> 2.45336318

12 13 c c 4.87596464

13 14 r r 6.10458231

14 15 o o 1.90881133

15 16 s s 4.37775946

16 17 e e 3.08709335

17 18 <space> <space> 3.18283558

18 19 o o 4.64191341

19 20 f f 2.57291579

20 21 <space> <space> 2.04530096

21 22 h h 5.0115099

22 23 a a 2.05648804

23 24 m m 6.21440411

24 25 e e 1.82324862

25 26 n n 4.63849545

26 27 <space> <space> 2.91914487

27 28 e e 5.65896082

28 29 v v 3.66322064

29 30 e e 1.19393516

30 31 n n 5.15902996

31 32 t t 3.26032686

32 33 s s 4.88752651

33 34 <space> <space> 2.14498758

34 35 i i 4.44708538

35 36 t t 3.59544587

36 37 <space> <space> 3.09999871

37 38 b b 5.45346117

38 39 e e 2.08525467

39 40 c c 5.96778488

40 41 a a 3.0632267

41 42 m m 4.40555239

42 43 s s 6.66972923

43 44 <space> <space> 2.51859236

44 45 n n 5.72867584

45 46 i i 3.21005249

46 47 c c 5.78328514

47 48 e e 3.03224134

48 49 s s 5.04981804

49 50 s s 3.71643329

50 51 e e 2.63232446

51 52 r r 6.06487942

52 53 y y 4.80678463

53 54 <space> <space> 2.27946734

54 55 f f 5.17545509

55 56 o o 2.2566762

56 57 r r 3.36134315

57 58 <space> <space> 3.02043962

58 59 o o 4.44632244

59 60 n n 3.81327963

60 61 e e 4.00009251

61 62 <space> <space> 3.15706658

62 63 p p 5.1651535

63 64 e e 2.65750647

64 65 o o 2.81087565

65 66 p p 4.25629711

66 67 l l 3.09632874

67 68 e e 2.71689034

68 69 <space> <space> 2.90927219

69 70 t t 4.19441509

70 71 o o 3.88887739

71 72 <space> <space> 2.45522833

72 73 d d 5.1143589

73 74 e e 2.52583623

74 75 s s 5.18131256

75 76 s s 3.71643329

76 77 e e 2.63232446

77 78 l l 5.38525581

78 79 v v 5.54194832

79 80 e e 2.42412567

80 81 <space> <space> 3.14656782

81 82 t t 4.19441509

82 83 h h 2.28522635

83 84 e e 2.35988283

84 85 <space> <space> 2.45336318

85 86 p p 5.1651535

86 87 l l 5.40753746

87 88 a a 2.04834318

88 89 t t 5.21706676

89 90 i i 3.22607231

90 91 c c 5.86377907

91 92 a a 2.97710586

92 93 l l 5.00540686

93 94 <space> <space> 3.56489825

94 95 b b 4.99757814

95 96 a a 3.55952811

96 97 n n 4.64118338

97 98 d d 2.54704547

98 99 s s 5.84989262

99 100 <space> <space> 13.5320959

**fstcompose declaration.fst squash.fst | fstproject --project\_output | fstrmepsilon | fstcompose – resquash5.fst |fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort |fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 4.60410023

1 2 h h 2.66454315

2 3 e e 2.35879779

3 4 n n 3.55122757

4 5 <space> <space> 2.04214239

5 6 i i 3.37982273

6 7 n n 3.07805848

7 8 <space> <space> 2.48634529

8 9 t t 2.80491948

9 10 h h 2.06189346

10 11 e e 2.15648317

11 12 <space> <space> 2.2382586

12 13 c c 4.45852709

13 14 u u 3.20814967

14 15 r r 4.05094528

15 16 s s 5.22375917

16 17 e e 2.89283442

17 18 <space> <space> 2.75336838

18 19 o o 4.91257477

19 20 f f 2.37501621

20 21 <space> <space> 2.02729487

21 22 h h 5.00897503

22 23 u u 2.68937397

23 24 m m 3.6347115

24 25 a a 1.20132816

25 26 n n 3.00035

26 27 <space> <space> 2.20361757

27 28 e e 6.16964579

28 29 v v 4.38147402

29 30 e e 1.11291671

30 31 n n 3.68742061

31 32 t t 4.21746635

32 33 s s 4.20740795

33 34 <space> <space> 2.00066233

34 35 i i 4.47609138

35 36 t t 3.40293694

36 37 <space> <space> 2.128407

37 38 b b 5.73617744

38 39 e e 2.51811337

39 40 c c 5.90978861

40 41 o o 2.31792307

41 42 m m 3.04758978

42 43 e e 1.22527313

43 44 s s 5.6012125

44 45 <space> <space> 2.01255655

45 46 n n 6.1854682

46 47 e e 1.64390683

47 48 c c 5.44074965

48 49 e e 1.43731618

49 50 s s 3.06878734

50 51 s s 2.28544641

51 52 a a 2.51453686

52 53 r r 3.78899503

53 54 y y 3.58357859

54 55 <space> <space> 2.01008129

55 56 f f 4.42279482

56 57 o o 2.16488862

57 58 r r 3.37866712

58 59 <space> <space> 2.32046366

59 60 o o 6.46300411

60 61 n n 3.57750273

61 62 e e 3.02500629

62 63 <space> <space> 2.06165981

63 64 p p 5.54689455

64 65 e e 2.74949217

65 66 o o 1.55567622

66 67 p p 4.00000906

67 68 l l 2.00009012

68 69 e e 2.00002003

69 70 <space> <space> 2.02203989

70 71 t t 4.43917561

71 72 o o 3.04503608

72 73 <space> <space> 2.42009258

73 74 d d 5.08002949

74 75 i i 2.13500357

75 76 s s 3.7211287

76 77 s s 5.60415268

77 78 o o 3.0116961

78 79 l l 3.5579617

79 80 v v 2.60895443

80 81 e e 2.00270414

81 82 <space> <space> 4.02176666

82 83 t t 3.72435093

83 84 h h 2.34322095

84 85 e e 2.55643368

85 86 <space> <space> 2.2382586

86 87 p p 4.67258644

87 88 o o 2.86510992

88 89 l l 5.25078917

89 90 i i 1.50907183

90 91 t t 5.28289318

91 92 i i 2.82918549

92 93 c c 4.88956642

93 94 a a 2.08810735

94 95 l l 3.21859312

95 96 <space> <space> 2.33520889

96 97 b b 4.48344707

97 98 e e 2.84059691

98 99 n n 7.89088058

99 100 d d 3.06457877

100 101 s s 6.59354877

101 102 <space> <space> 17.7748604

**fstcompose declaration.fst squash.fst | fstproject --project\_output | fstrmepsilon | fstcompose – resquash7.fst |fstshortestpath | fstproject --project\_output | fstrmepsilon | fsttopsort |fstprint --isymbols=ascii.syms --osymbols=ascii.syms**

0 1 w w 4.60410023

1 2 h h 2.66454315

2 3 e e 2.35879779

3 4 n n 3.55122757

4 5 <space> <space> 2.00591278

5 6 i i 2.75286913

6 7 n n 7.09879732

7 8 <space> <space> 2.17637062

8 9 t t 3.1464448

9 10 h h 2.06371593

10 11 e e 2.08522701

11 12 <space> <space> 2.07994556

12 13 c c 4.41605091

13 14 o o 2.13718605

14 15 u u 3.07013893

15 16 r r 6.06479836

16 17 s s 2.82562995

17 18 e e 2.00237775

18 19 <space> <space> 2.11576366

19 20 o o 4.01724529

20 21 f f 2.03536487

21 22 <space> <space> 2.0073452

22 23 h h 4.32929659

23 24 u u 2.39522171

24 25 m m 3.56074691

25 26 a a 1.00107956

26 27 n n 3.00000048

27 28 <space> <space> 2.21061182

28 29 e e 5.64718723

29 30 v v 5.76776838

30 31 e e 1.22597039

31 32 n n 3.60241222

32 33 t t 6.29850483

33 34 s s 2.43687916

34 35 <space> <space> 2.00002646

35 36 i i 5.37661839

36 37 t t 2.87636948

37 38 <space> <space> 2.14175606

38 39 b b 6.4954071

39 40 e e 2.39380074

40 41 c c 7.45023346

41 42 o o 3.92736101

42 43 m m 3.00290895

43 44 e e 1.14419329

44 45 s s 10.8793268

45 46 <space> <space> 2.00178432

46 47 n n 8.48805428

47 48 e e 1.27099776

48 49 c c 3.57436132

49 50 e e 2.65568495

50 51 s s 3.00302601

51 52 s s 2.00205421

52 53 a a 1.70883012

53 54 r r 3.07062626

54 55 y y 2.56603074

55 56 <space> <space> 2.00062704

56 57 f f 3.30133438

57 58 o o 1.33036757

58 59 r r 3.34696341

59 60 <space> <space> 2.262146

60 61 o o 8.117383

61 62 n n 2.05086637

62 63 e e 2.11323619

63 64 <space> <space> 2.02099943

64 65 p p 4.19642735

65 66 e e 2.88093567

66 67 o o 1.65696752

67 68 p p 4

68 69 l l 2

69 70 e e 2

70 71 <space> <space> 2.01904869

71 72 t t 4.5633707

72 73 o o 3.13061786

73 74 <space> <space> 2.21024466

74 75 d d 4.20483541

75 76 i i 2.11217284

76 77 s s 3.7410965

77 78 s s 8.16910076

78 79 o o 2.74205208

79 80 l l 3.11306381

80 81 v v 2.12106109

81 82 e e 2.0006752

82 83 <space> <space> 4.07286644

83 84 t t 2.86851978

84 85 h h 3.72951531

85 86 e e 2.40056944

86 87 <space> <space> 2.37881565

87 88 p p 4.71044064

88 89 o o 2.86137486

89 90 l l 4.99965429

90 91 i i 1.18670535

91 92 t t 5.03669834

92 93 i i 1.23555017

93 94 c c 3.41130567

94 95 a a 1.18126607

95 96 l l 3.0503335

96 97 <space> <space> 2.21470952

97 98 b b 3.76300621

98 99 e e 4.96600103

99 100 n n 7.89088058

100 101 d d 3.06457877

101 102 s s 8.09762573

102 103 <space> <space> 17.7748604

Q.3)

Output Symbol Table:

Number.txt

0 0

1 1

2 2

3 3

4 4

5 5

6 6

7 7

8 8

9 9

one 10

two 11

three 12

four 13

five 14

six 15

seven 16

eight 17

nine 18

ten 19

eleven 20

twelve 21

thirteen 22

fourteen 23

fifteen 24

sixteen 25

seventeen 26

eighteen 27

nineteen 28

twenty 29

thirty 30

forty 31

fifty 32

sixty 33

seventy 34

eighty 35

ninety 36

hundred 37

thousand 38

million 39

<epsilon> 40

zero 41

1. Compile the Finite State Transducer consisting of 21 states using the output symbol table number.txt

**fstcompile --isymbols=ascii.syms --osymbols=number.txt < demo3.txt > demo3.fst**

**Demo3.txt**

0 20 <epsilon> <epsilon>

0 18 <epsilon> <epsilon>

0 15 <epsilon> <epsilon>

0 13 <epsilon> <epsilon>

0 11 <epsilon> <epsilon>

0 8 <epsilon> <epsilon>

0 6 <epsilon> <epsilon>

0 2 <epsilon> <epsilon>

0 1 <epsilon> <epsilon>

1 21 0 zero

1 21 1 one

1 21 2 two

1 21 3 three

1 21 4 four

1 21 5 five

1 21 6 six

1 21 7 seven

1 21 8 eight

1 21 9 nine

2 3 0 <epsilon>

2 3 2 twenty

2 3 3 thirty

2 3 4 forty

2 3 5 fifty

2 3 6 sixty

2 3 7 seventy

2 3 8 eighty

2 3 9 ninety

3 21 0 <epsilon>

3 21 1 one

3 21 2 two

3 21 3 three

3 21 4 four

3 21 5 five

3 21 6 six

3 21 7 seven

3 21 8 eight

3 21 9 nine

2 4 1 <epsilon>

4 21 0 ten

4 21 1 eleven

4 21 2 twelve

4 21 3 thirteen

4 21 4 fourteen

4 21 5 fifteen

4 21 6 sixteen

4 21 7 seventeen

4 21 8 eighteen

4 21 9 nineteen

6 2 0 <epsilon>

6 5 1 one

6 5 2 two

6 5 3 three

6 5 4 four

6 5 5 five

6 5 6 six

6 5 7 seven

6 5 8 eight

6 5 9 nine

5 2 <epsilon> hundred

7 6 <epsilon> thousand

8 7 1 one

8 7 2 two

8 7 3 three

8 7 4 four

8 7 5 five

8 7 6 six

8 7 7 seven

8 7 8 eight

8 7 9 nine

11 9 0 <epsilon>

11 9 2 twenty

11 9 3 thirty

11 9 4 forty

11 9 5 fifty

11 9 6 sixty

11 9 7 seventy

11 9 8 eighty

11 9 9 ninety

9 7 0 <epsilon>

9 7 1 one

9 7 2 two

9 7 3 three

9 7 4 four

9 7 5 five

9 7 6 six

9 7 7 seven

9 7 8 eight

9 7 9 nine

11 10 1 <epsilon>

10 7 0 ten

10 7 1 eleven

10 7 2 twelve

10 7 3 thirteen

10 7 4 fourteen

10 7 5 fifteen

10 7 6 sixteen

10 7 7 seventeen

10 7 8 eighteen

10 7 9 nineteen

13 12 1 one

13 12 2 two

13 12 3 three

13 12 4 four

13 12 5 five

13 12 6 six

13 12 7 seven

13 12 8 eight

13 12 9 nine

13 11 0 <epsilon>

12 11 <epsilon> hundred

14 13 <epsilon> million

15 14 1 one

15 14 2 two

15 14 3 three

15 14 4 four

15 14 5 five

15 14 6 six

15 14 7 seven

15 14 8 eight

15 14 9 nine

18 16 0 <epsilon>

18 16 2 twenty

18 16 3 thirty

18 16 4 forty

18 16 5 fifty

18 16 6 sixty

18 16 7 seventy

18 16 8 eighty

18 16 9 ninety

16 14 0 <epsilon>

16 14 1 one

16 14 2 two

16 14 3 three

16 14 4 four

16 14 5 five

16 14 6 six

16 14 7 seven

16 14 8 eight

16 14 9 nine

18 17 1 <epsilon>

17 14 0 ten

17 14 1 eleven

17 14 2 twelve

17 14 3 thirteen

17 14 4 fourteen

17 14 5 fifteen

17 14 6 sixteen

17 14 7 seventeen

17 14 8 eighteen

17 14 9 nineteen

20 19 1 one

20 19 2 two

20 19 3 three

20 19 4 four

20 19 5 five

20 19 6 six

20 19 7 seven

20 19 8 eight

20 19 9 nine

19 18 <epsilon> hundred

21

**Sample Test Case:**

1. **Compile the input Acceptor**

**fstcompile --isymbols=ascii.syms -acceptor input.txt input.fst**

Input.txt

0 1 1

1 2 1

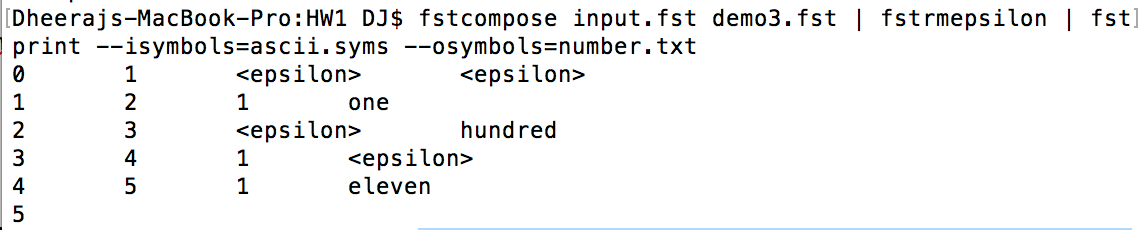
2 3 1

3

1. **Compose with the Demo3.fst**

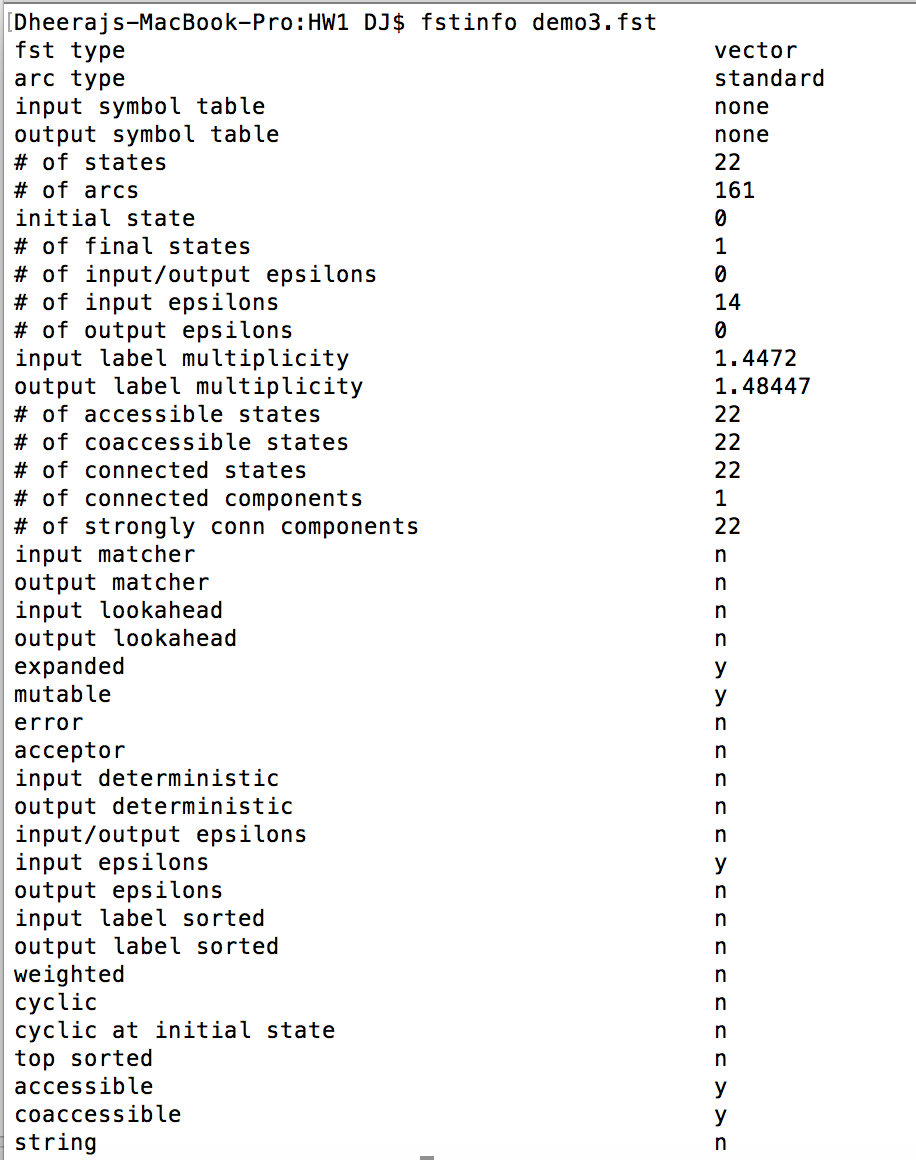
**fstcompose input.fst demo3.fst | fstrmepsilon | fstprint --isymbols=ascii.syms --osymbols=number.txt**

**Output:**



**FST Info:**

fstinfo demo3.fst

****

**FST Diagram:**

**fstdraw --isymbols=ascii.syms --osymbols=number.txt demo3.fst demo3.dot**

**dot -Tps demo3.dot > demo3.ps**

