Assignment 2

1.

a.

Given the FSA, the accepted strings are:

* Ab
* Abab
* ababab
* Aab
* Aababaab

Look at the accepted strings, the language must end with the term ab or aab and starts with a.

Hence, the regular expression is (ab + aab) (ab + aab)\*

b.

Given the NFSA, the accepted strings are:

* Ab
* Abab
* Aba
* Abaab
* Abaaba
* Abaabab
* Abaabaaba

Look at the accepted strings, the language must end with the term ab or ba and starts with a.

* Hence, the regular expression is (ab + aba) (ab + aba)\*

2. Compute bigram probabilities

a.

Total words = 16

Unigram count and probability

|  |  |  |
| --- | --- | --- |
| Unigram | Count | Probability |
| a | 6 | 6/16 = 3/8 |
| <s> | 1 | 1/16 |
| </s> | 1 | 1/16 |
| man | 3 | 3/16 |
| plan | 2 | 1/8 |
| canal | 1 | 1/16 |
| panama | 2 | 1/8 |

Bigram count and probability

|  |  |  |
| --- | --- | --- |
| Bigram | Count | Probability |
| [<s> a] | 1 | 1 |
| [ a man] | 3 | ½ |
| [man a] | 3 | 1 |
| [a plan] | 2 | 1/3 |
| [plan a] | 2 | 1 |
| [a canal] | 1 | 1/6 |
| [canal panama] | 1 | 1 |
| [panama panama] | 1 | ½ |
| [panama <s>] | 1 | 1/2 |

b.

1. Compute the bigram-based probability without smoothing for the following sentence

*<s> plan a panama </s>*

P(sentence) = P(<s>)\*P(plan|<s>)\*P(a|plan)\*P(panama|a)\*P(<s>|panama)

Because the birgram <s> plan does not exist in the training corpus, the p(plan | <s>) = 0

Hence, the P(sentence) = 0

1. Compute the bigram-based probability with add-one-smoothing for the following sentence

*<s> plan a panama </s>*

Bigram count and probability with add-one-smoothing

|  |  |  |
| --- | --- | --- |
| Bigram count w/ add-one-smoothing | Count | Probability |
| [<s> a] | 2 | 2/(7 + 1) = 1/4 |
| [a man] | 4 | 4/(7 + 6) = 4/13 |
| [man a] | 4 | 4/(7 + 3) = 1/5 |
| [a plan] | 3 | 3/(7 + 6) = 3/13 |
| [plan a] | 3 | 3/(7 + 2) = 1/3 |
| [a canal] | 2 | 2/(7 + 6) = 2/13 |
| [canal panama] | 2 | 2/(7 + 1) = 1/4 |
| [panama panama] | 2 | 2/(7 + 2) = 2/9 |
| [panama </s>] | 2 | 2/(7 + 1) = 1/4 |
| [<s> plan] | 1 | 1/(7 + 1) = 1/8 |
| [a panama] | 1 | 1/(7 + 6) = 1/13 |

P(sentence) = P(<s>)\*P(plan|<s>)\*P(a|plan)\*P(panama|a)\*P(<s>|panama)

= (1/16)\*(1/8)\*(1/3)\*(1/13)\*(1/4) = 0.00005008

1. Compute the bigram-based probability with good-turing smoothing for the following sentence

*<s> plan a panama </s>*

Bigram count and probability of bigrams in the sentence

|  |  |  |
| --- | --- | --- |
| Bigram | Count | Probability |
| [<s> plan] | 0 | N1/N0 = 5/16 |
| [plan a] | 2 | C\*/N = (3\*2/2)/16 = 3/16 |
| [a panama] | 0 | N1/N0 = 5/16 |
| [panama </s>] | 1 | C\*/N = (2\*2/5)/16 = 1 /20 |

P(sentence) = P(<s>)\*P(plan|<s>)\*P(a|plan)\*P(panama|a)\*P(<s>|panama)

= (1/16) \* (5/16)\*(3/16)\*(5/16)\*(1/20) = 75/(1310720) = 0.00005722