Assignment 2

Writers’s aids

Due Friday, September 23

1. Do question #1a in chapter 2 of the draft textbook (p. 79).

2. Pretend we have a bigram array, as in the given table, where the first letter of the bigram is given in the vertical letters (i.e., down the side), and the second letter is given in the horizontal ones (i.e., across the top)

<table>
<thead>
<tr>
<th>i</th>
<th>j</th>
<th>k</th>
</tr>
</thead>
</table>
   i |   |   |   |
   j |   |   |   |
   k |   |   |   |

Make this into 3 positional bigram arrays, namely ones which capture the positions:

- start of word
- middle of word
- end of word

For the “middle of word” one, provide a word which justifies each 1 you put in the chart.

3. Do question #3 in chapter 2 of the draft textbook (p. 80).

   • **Bonus:** Do question #4 in chapter 2 of the draft textbook (p. 80).

4. Consider the misspelling forg, and assume our edit distance calculations have insertions, deletions, substitutions, and transpositions. Describe how probabilities are used to rank frog vs. forge. Do you have an intuition as to which should be higher, and why?

5. Pick a search engine of your choice, as long as the search engine provides spelling suggestions. Search with the keywords ghouls and ghosts, and do not use quotes.

   (a) Try as many sensible misspellings as you can think of (and vary the order of the words, too). Which ones does the search engine catch and which doesn’t it catch?

   (b) Given what we talked about in class, can you explain why certain variations were flagged as misspellings and others were not? If the search engine always correctly caught your misspellings, it will help to try some even more varied misspellings, in order to find the “boundary” where the correct suggestion is no longer made.

6. Which method for correcting spelling errors do you hypothesize would work best for an Optical Character Recognition (OCR) system? for a writer’s errors? for errors made by a non-native speaker? Why would the method be good for that task? (Hint: Think about the different types of errors that would be common in each case.)
7. Look at the following examples of mistakes made by smartphone auto-correct programs.

- “Sitting at the dentist waiting for my son...they said he has NO favorites!”
  cavities → favorites
- “I’m like $300 short on my M OTTSAPPLESAUCE payment due the 15th.”
  mortgage payment → M OTTSAPPLESAUCE
- “I’ll get back with you by tomorrow. We’re planning my starvation”
  staycation → starvation
- “Thanks for letting me centipede to you today by the way”
  vent → centipede
- “I have to go get my kids. my parrots watched them last night”
  parents → parrots
- “Please bring anvils if you can”
  a coke → anvils
- “Sry I was teething”
  tweeting → teething
- “don’t worry, you look super affordable”
  adorable → affordable
- “T.G.I.Fibromyalgia”
  FRIDAYYYYY → Fibromyalgia

(a) Try typing a few misspellings for each of the original words in Microsoft Word, ispell/aspell, or a similar word processing spelling checker. Are the suggested corrections similar to the ones made by the smartphone auto-correct program?

(b) What do you think might be happening to cause the smartphones to make such different corrections?

---

1 Adapted from http://www.facebook.com wall post.
2 http://damnyouautocorrect.com/images/applesauce.jpg
3 http://static.businessinsider.com/image/4c96f684bd7c885630b0000-590/not-starvation.jpg
4 http://i.huffpost.com/gadgets/slideshows/13111/slideshow_13111_178355_large.jpg?1315764213428
5 http://i.huffpost.com/gadgets/slideshows/13111/slideshow_13111_178372_large.jpg?1315764398048
6 http://i.huffpost.com/gadgets/slideshows/13111/slideshow_13111_178377_large.jpg?1315764384159
8 http://i.huffpost.com/gadgets/slideshows/13111/slideshow_13111_179365_large.jpg?1315763724269
9 http://damnyouautocorrect.com/images/tgifi.jpg