T9 and Tries

CSE 303 Homework 4, Winter 2009
T9 Predictive Text

• **What is T9? Demo**

• **T9onyms:**
  1. **22737:** acres, bards, barer, bares, baser, bases, caper, capes, cards, cares, cases
  2. **46637:** goner, goods, goofs, homer, homes, honer, hones, hoods, hoofs, inner
  3. **2273:** acre, bard, bare, base, cape, card, care, case
  4. **729:** paw, pay, Paz, raw, ray, saw, sax, say
  5. **76737:** pores, poser, poses, roper, ropes, roses, sorer, sores

• **How does T9 order T9onyms?**
  – Assignment Requirement: Alphabetical order
  – Extra credit options: Frequency, Dynamic Frequency
Trie

• Tree structure: n-ary tree
• We use a trie to store pieces of data that have a key (used to identify the data) from an alphabet
  – Optionally can also hold a value (which holds any additional data associated with the key).

• Applications:
  – Spell checkers
  – Auto-complete
  – Data compression
  – T9 predictive text input for cell phones
  – String search
Example: String Search

• Goal:
  – Determine if a given word appears in a block of text.
  – Optimize for multiple searches in the same block of text

• What do we do?
  – Place each word in the block of text into a data structure
  – Use data structure to determine whether a word exists in that block of text

• Which data structure should we use?
String Search Trie

- Text: sells sea shells by the shore
String Search Trie

- Search for: shells
Building a Trie for T9

• How is a T9 Trie different?
  – Alphabet: \{2-9\}
Handling T9onyms
Handling T9onyms
Extra Credit

• More accurately implement T9:

1. Store the prefix of each word in the text file in the trie—
   - example: foobar - “f”, “fo”, “foo”, “foob”, “fooba”, “foobar”

2. Order a word in the Trie by its frequency
   - A word with a higher frequency will be predicted before a lower frequency word. In the text file (listed on assignment), the format of an entry will be

   \[
   \text{word} \quad \text{frequency}
   \]

3. Update the frequency of a word
   - Each time a word is used, increment the word’s frequency
   - Prediction of words should use updated frequencies