

### 3. (10 points) F u cn rd ths

Abbreviations are hard. We are used to thinking of standard abbreviations like km, UK, Mr or Ave. But in fact people make up new abbreviations all the time, if they are under time pressure (e.g. instant messaging) or if they have severe space limitations (e.g. classified ads in a printed newspaper).

One place where you find lots of abbreviations is the notes taken by the overworked people who staff call centres. They have to record what was discussed, but they don't have the time to type everything out. So you often get things that look like this, from the logs of a call centre run by a major telecommunications company:

cust rcvd lttr cncrng local srvc

which of course is supposed to mean:

customer received letter concerning local service

Let's say you are designing a computer program to try to flesh out abbreviated messages automatically. You can't just have a fixed list of abbreviations: the set is pretty open ended. But what the computer can do is to look at all the messages it's got, and hope that someone somewhere has spelled out the complete words. So if for example it is looking at *rcvd lttr*, and somewhere else in the database someone has done it the favour of reporting on a different call using the fully spelled phrase *received letter*, then it has a chance of guessing the expansion of *rcvd lttr*. That is, *rcvd* is a plausible abbreviation of *received*, *lttr* is a plausible abbreviation of *letter*, and the two occur together in the right order.

Of course, you know English, so you could have figured this out. But the computer really doesn't. To the computer the problem looks as follows: It has a bunch of words and phrases, some of which are abbreviated, written in a bunch of otherwise meaningless symbols (remember the computer doesn't know English and to it, the strings are ultimately just a bunch of characters anyway).

Here are some messages (A-R), transposed into symbols so that you see what the computer "sees", and below them a list of full phrases from elsewhere in the collection (1-15). There are two warnings to bear in mind:

- (1) When you are under time pressure, you make mistakes. There are actually three typos in the abbreviations—typos in that all the letters are there, but are out of the expected order, and therefore are not strictly speaking reasonable abbreviations for the words.
- (2) There are three phrases in the second list that are not found in the abbreviations, and (of course), any of these phrases may be represented by more than one of the abbreviations.

- A.  $\neg \theta \odot \oplus \cap \sqcup$
- B.  $\neg \odot \neg \pm \circ \circ \cap \times$
- C.  $\neg \theta \emptyset \bullet \oplus \pm \times \circ \theta \times$
- D.  $\neg \emptyset \theta \neg \pm \circ \circ \cap \times$
- E.  $\neg \odot \theta \pm \times \circ * \theta \cap \times$
- F.  $\neg \wedge \bullet \odot \cap \theta \cap \emptyset$
- G.  $\neg \theta \emptyset \oplus \neg \circ \pm * \bullet \theta$
- H.  $\neg \odot \theta \emptyset \bullet \neg \circ \circ \times$
- I.  $\neg \bullet \oplus \times \theta \neg \vee \vee \neg \emptyset \times$
- J.  $\neg \odot \theta \emptyset \oplus \odot \cap \theta \emptyset$
- K.  $\neg \theta \emptyset \odot \vee \times \cap \theta \emptyset \emptyset \wedge \wedge \times$
- L.  $\neg \theta \odot \emptyset \ddagger \vee \emptyset \times$
- M.  $\neg \theta \emptyset \bullet \ddagger \vee \emptyset \theta$
- N.  $\neg \theta \neg \circ \cup$
- O.  $\neg \theta \oplus \neg \pm \circ \circ \cap \times$

P.  $\overline{F}\ominus\bullet\oplus$   $\overline{F}\circ\circ\vee U$

Q.  $\overline{F}\odot\ominus\oslash$   $\overline{F}\pm\oplus\cap$

R.  $\overline{F}\odot\ominus\oslash\wedge$   $\overline{F}\pm\circ\circ$

- (1) customer advised
- (2) customer call
- (3) customer called
- (4) customer calling
- (5) customer care
- (6) customer claims
- (7) customer disconnected
- (8) customer likes
- (9) customer needs
- (10) customer request
- (11) customer says
- (12) customer understood
- (13) customer upset
- (14) customer wanted
- (15) customer wants

**3.1. (7 points)** In the table on the answer sheet, match each of the encoded abbreviations from the previous page to one of the phrases above.

3.2. (3 points) What message is abbreviated in the symbols below?

F ⊖ ⊘ ⊖ ⊕ ± \* × ‡ √ ⊖ ⊖ ⊖ ∆

× ⊕ ∏ ⊖ ⊙ ∓ ∓ ⊙ ⊖ ⊖ ∆

∓ ∘ ∘ × ∓ ∘ ± \* ● ⊖ × \* × √ ⊖

⊕ ∓ ⊙ ∘ ⊖ ⊖ ⊕ ∓ ∓ √ ∪ ⊕ ⊖ ⊖

⊙ ∪ ⊖ ⊖ ∪ ∩ ⊖ ∆ ∘ ⊙ ⊖ \* ∆ √

|       |   |   |   |   |
|-------|---|---|---|---|
| 3.1 A | B | C | D | 7 |
| E     | F | G | H |   |
| I     | J | K | L |   |
| M     | N | O | P |   |
| Q     | R |   |   |   |
| 3.2   |   |   |   |   |

### Solutions

|   |             |             |             |          |
|---|-------------|-------------|-------------|----------|
| <b>3.1 A 10</b>   | <b>B 3</b>  | <b>C 1</b>  | <b>D 3</b>  | <b>7</b> |
| <b>E 1</b>  | <b>F 13</b> | <b>G 6</b>  | <b>H 3</b>  |          |
| <b>I 7</b>  | <b>J 13</b> | <b>K 12</b> | <b>L 14</b> |          |
| <b>M 15</b>   | <b>N 4</b>  | <b>O 3</b>  | <b>P 4</b>  |          |
| <b>Q 5</b>  | <b>R 2</b>  |             |             |          |
| <b>3.2</b><br>cst said wnts to drp svc<br>custo cld claims didnt rev ltr<br>cngrts u gt the solution<br>(Customer said wants to drop service<br>Customer called claims didnt receive letter<br>Congrats you got the solution) |             |             |             | <b>3</b> |