School of Computing and Information Systems
The University of Melbourne
COMP90049
Knowledge Technologies (Semester 1, 2018)
Workshop exercises: Week 3

1. Following on from last week, write a **regular expression** which will:

   (a) Match a string according to whether it contains a price (like \$20 or \$0.99, but not 11.30 or 0\$n1a).

   (b) Match a number in scientific E notation (e.g. `2.00600e+003`)

   (c) Remove all HTML comments from an HTML document (defined as a string)

   (d) Validate an email address (i.e. the string will match if it is an email address, and will mismatch otherwise)

———————————————————— & ————————————————————

Suppose that we have observed the token `lended`, and we have a dictionary as follows:

```
addendum
blenders
commodity
deaden
end
leader
leant
lent
lemonade
pleading
```

2. Which, if any, of the above dictionary entries be returned using a Neighbourhood Search with a neighbourhood of 1? 2? 3?

3. With respect to the input string `lended` and the dictionary entry `deaden`, calculate the following:

   (a) the Global Edit Distance, using the parameter $[m, i, d, r] = [+1, -1, -1, -1]$

4. Find the best approximate match (or matches, if there are ties) in the dictionary for the string `lended`, based on the following methods; consider different parameters where necessary:

   (a) the Global Edit Distance

   (b) (continued next week)