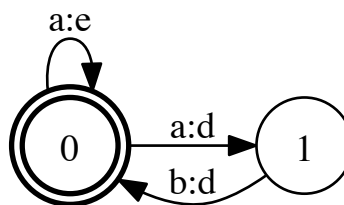


Goals of this lab:

- Make students understand how to use Regular Expressions.
- Practice the use of some evaluation metrics.
- Build a small chatbot based on Regular Expressions and understand how difficult it is to build a chatbot that is at the human level.

1. Consider the transducer (uhhh, do you still remember them?).



Choose the only correct option. ALL the possible output sequences of this transducer can be represented with the following regular expression:

1. $(e^*(dd)?e^*)^*$
2. $(e^*[dd]e^*)^*$
3. $(e^+(dd)?e^+)^*$
4. $(e^+[dd]e^*)^*$
5. $(e^+(dd)e^*)^*$

2. Write, in cell (2, 2), the sequences that respect (both) Regular Expressions from cells (1, 2) and (2, 1).

	[^AB]
[ABC]	

3. Write a regular expression defined with the words (total sequence, do not split them into characters) “hello” and “bye”. Define a regular expression that (only) recognizes:

3.1 The set of sequences that end in “byebye”

3.2 The set of sequences where each pair of “hello”s is followed by a pair of “bye”s

4. Fill in the table, one letter in each cell, considering that the regular expressions indicate the letters that should appear in each column/line. As an example, in the first line you might have an H in the first cell and an E in the second, or an L in each cell or, at least, an O in one of the cells:

	[rSPEAK]+	EPIPIEF
HE LL O+		
[PLEASE]+		

5. Consider systems A and B, which target the extraction of named entities of type **Location** from text. Considering a given Golden Collection, and A and B results (Table below), find Precision, Recall and F-measure (F_1) for both systems.

Golden Collection	System A	System B
Igreja de São Domingos, Rua Augusta, Praça do Comércio, Rua do Arsenal, Santa Engrácia	Igreja de São Domingos, Praça do Comércio, Rua do Arsenal, Santa Engrácia, S. Julião, KnitPro	Rua Augusta, Praça do Comércio, Rua do Arsenal

	System A	System B	Formulas
True Positives			Precision: $P = TP / (TP + FP)$
False Positives			
False Negatives			Recall: $R = TP / (TP + FN)$
Precision			
Recall			$F_1: 2PR / (P + R)$
F_1			

6. Create your own pattern-based chatbot based on the ELIZA implementation present in NLTK (the code needed is given to you).

6.1 Interact with ELIZA. To run your code, use:

```
python3 Eliza-nltk.py
```

6.2. Imagine that ELIZA is now a nutritionist that wants to be sure that you have a healthy diet. Use the given files to create the new ELIZA. Play with the possibility of replacing words or sequences of words. Have fun.

6.3. How do you think that you could systematically develop and evaluate such system?