

LÍNGUA NATURAL 2020/2021

Mini-Project Nº 2 (MP2)

Should be done:

☐

individually

☒

in group

Submission:

☐

theoretical class

☒

Fenix submission

Submission deadline: till 15:59, 16/Nov

OBJECTIVES

Learn to work with transducers, using them to solve a problem.

STATEMENT

Suppose you want to create a time conversion module, using only transducers, to couple with a speech recognition system. To do this, the module must convert an hour that is written in full into the corresponding numerical format **hh:mm**. Some examples:

- *oito horas e quinze minutos* → 08:15.
- *oito e quinze* → 08:15;
- *catorze e vinte cinco* → 14:25;
- *dez horas e um quarto* → 10:15.

1. In order to produce this module, develop the following transducers:

- a. the transducer **horas** converts hours written in full to the 2-digit numeric format, considering numbers from the [0..23] range and the *uma* and *duas* variants. The word *hora(s)* is optional. Examples: *uma* is converted to 01, *vinte e duas horas* is converted to 22;
- b. the transducer **minutos** similar to the previous one, but considering numbers from the [0..59] range and considering the *um* and *dois* variants and. The word *minuto(s)* is optional. Examples: *seis minutos* is converted to 06, *vinte e um* is converted to 21;
- c. the transducer **meias** converts the expression *meia* into an equivalent expression using numbers. Example: *meia* is converted to *trinta*;
- d. the transducer **quartos** converts expressions using *quarto* and *quartos* into the equivalent expression using numbers. Examples: *um quarto* is converted into *quinze* and *dois quartos* is converted into *trinta*.

2. Using the previous transducers (the previous 4 transducers cannot be modified) and, if needed, other transducers, develop:

- e. transducer **text2num**, which converts natural language normalized expressions that specify hours ("X [*horas*] e Y [*minutos*]") into the numerical condensed form **hh:mm**. Examples: *dez horas e quinze*, *dez e quinze minutos*, *dez horas e quinze minutos* are converted into 10:15. *dez e zero* is converted into 10:00;
- f. transducer **lazy2num**, which being a variant of the previous one, accepts simpler expressions in

the form "X [horas]". Examples: *dez* is converted to *10:00*. This transducer must also accept all the expressions accepted by `text2num.fst`;

- g. transducer **rich2text**, which converts expressions that specify hours using *meia* and *quarto(s)* into the equivalent expression without those words. Examples: *dez e um quarto* is converted to *dez e quinze*, *dezoito e tres quartos* is converted to *dezoito e quarenta e cinco*;
- h. transducer **rich2num**, which converts any expression that specifies hours into its condensed numeric form **hh:mm**. Examples: *dez* is converted to *10:00*, *dez e meia* is converted to *10:30*, *dez e quinze* and *dez e um quarto* are converted to *10:15*;
- i. transducer **num2text**, which converts any hour in its condensed numerical form to the corresponding simplified text expressions always using the words *hora(s)* and *minuto(s)*. Example: *10:15* is converted to *dez horas e quinze minutos*.

3. Test the transducers `rich2num` and `num2text` with the hours at which each member of the group gets up and goes to bed on Saturdays. Knowing that the transducers to test do not accept the same input, you should designate each of your tests as `sleepA_XXXXX` and `wakeupA_XXXXX`, where A represents a letter from A to Z, corresponding to each of your tests. For example, `sleepD_12345.txt`, `sleepD_12345.fsm`, `sleepD_12345.pdf`.

Assume that:

- the file "syms.txt" has the symbols to be manipulated by the transducers and cannot be changed. Note that the symbols do not include words with diacritics and may contain symbols in addition to those required;
- the base transducers (task 1) cannot make conversions beyond what is required;
- you can use other transducers not mentioned in the statement;
- do not consider expressions like *um quarto para as nove*, *duas da tarde*, *meia noite* and *meio dia*.

SOFTWARE

To test the proposed solution use, in Linux environment, the tools:

- "OpenFST" da Google (<http://www.openfst.org/twiki/bin/view/FST/FstDownload>).
- "Graphviz" (<http://www.graphviz.org/>);

SUBMISSION

Submit in Fenix, project *MP2*, a zip file with, and only with:

- a shell script [the name has to be "run.sh"] with **all** the commands used to generate all transducers, either in binary and in graphical format (PDF, PS or PNG) from the ".txt" files;
- a folder "sources" containing all the text files used to define the transducers (extension ".txt");
- a folder "tests" with all the source test files (extensions ".txt");
- a folder "compiled" containing all the compiled version of all the transducers used, including the tests (extension ".fst");
- a folder "images" containing the graphical versions of all the transducers, including the tests (extension ".pdf", ".ps" or ".png");
- a short report, whose file name should be "report.txt" or "report.pdf", with a maximum of 1 page, containing the identification of the members of the group, the description of the options taken and comments on the solution developed. The report ALWAYS provide an estimate of each element's contribution to the work. For example: Peter: 60%, John: 40%, together with a small justification.

You can make several submissions: a new submission replaces the previous one.

Attention:

- developed transducers must have exactly the same names as above;
- the 4 folders "sources", "tests", "compiled" and "images" should not contain sub-folders.

EVALUATION CRITERIA

The following criteria will be taken into account in the assessment (maximum = 20 points):

1. Use of unnecessary writing of transducers (up to -4 points);
2. Correct operation of the first 4 transducers (1 point each);
3. Correct operation of the last 5 transducers (2 points each);
4. Run.sh operating correctly (3 points);
5. Delivery of the graphic versions of all transducers, as well as the examples, in their different forms, that is, after passing through the transducers (2 points);
6. Quality of the report [in Portuguese or in English] including spelling and syntactic correction (1 point);

Non-compliance with any rule implies a minimum discount of 4 points (in 20 points).

"POSSIBILITIES FOR PROMOTING ACADEMIC INTEGRITY" IN CARNEGIE MELLON UNIVERSITY

Both instructors and students can consider steps to enhance academic integrity in the CMU community. This section offers suggestions drawn from ongoing conversations with CMU students and faculty over the years and from the literature on academic integrity. The steps below include ways students can more effectively manage their own learning with the help of university resources and ways individual instructors can enhance support for student learning and integrity.

Steps Students Might Take:

- Ask about policies regarding collaboration and citations at the beginning of each course. Instructors' policies may differ substantially from one another.
- Ask questions - in class, immediately after class, in e-mail or in office hours - about course content or course procedures. If you are confused, you might ask for more clarification, different examples, or specific applications to help you understand. Other students often have the same questions you do so your questions can enhance the overall effectiveness of the course.
- Find out whether the instructor will provide suggestions for preparing for exams and consider preparing your own review sheet. The process of making a review sheet is actually a good method of improving your understanding of and memory for complex information.
- Refine your note-taking skills. Many students form the habit of transcribing whatever the professor writes, no more and no less. To facilitate better review and study sessions, ask yourself frequent questions as you read or listen to a lecture: What is the key new idea here? How can I use this information? Can I anticipate what is coming next?
- Improve your time management, especially during the day and early evening. Procrastination more often leads to ineffective cramming and loss of sleep than to good performance under pressure. If you begin to work well before due dates and examinations, you are much more likely to learn the material, to be able to get help if you need it, to feel less stressed, to perform better, and to avoid poor decisions on very late nights.
- Speak with the professors about their grading and homework policies if you feel that the policies

seem unfair - feedback is essential to improving the quality of a class. If you feel uncomfortable talking with an instructor directly, you might express your views in early course evaluations or to a teaching assistant.