LÍNGUA NATURAL 2021/2022

Mini-Project Nº 1 (MP1)

Should be done:	☐ individually	☒ **in group**
Submission:	☐ theoretical class	☒ **Fenix submission**
Submission deadline:	**till 23:59, 20/October**	

OBJECTIVES

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

STATEMENT

Suppose you want to convert "Roman birthdates" into an "Arabic-text birthdate" format. Some examples:

- *VIII/IX/MMXIII* → *08/Sep/2013*.
- *I/I/M* → *01/Jan/1000*.
- *X/XII/III* → *10/Dec/0003*.

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

    a. the transducer **mm2mmm** converts from 2-digit Arabic numbers to 3-letter (english) month names (e.g., *06* → *Jun*);

    b. the transducer **d2dd** converts Arabic numbers with a single digit to 2-digit numbers (e.g., *8* → *08*). Numbers with 2 or more digits are unchanged;

    c. the transducer **d2dddd** converts Arabic numbers with 1, 2, or 3-digits to 4-digit numbers (e.g., *8* → *0008*, *62* → *0062*, *753* → *0753*). Numbers with 4 or more digits are unchanged.

    d. the transducer **copy** only accepts a single symbol: a digit or a "/" (e.g., *0* → *0*, *1* → *1*, *8* → *8*, *9* → *9*, */* → */*) and fails if the input consists of any other symbol;

    e. the transducer **skip** converts a single symbol: a digit or a "/" into "eps" (e.g., *0* → *eps*, *1* → *eps*, *8* → *eps*, *9* → *eps*, */* → *eps*) and fails if the input consists of any other symbol;

    f. the transducer **date2year** selects the year in *Arabic birthdates* (e.g., *08/09/2013* → *2013*);

    g. the transducer **leap** analyses whether one year (4-digit numbers between 1901 and 2099) is a "not-leap" or a "leap" year (e.g., *1901* → *not-leap*, *1904* → *leap*);

    h. the transducer **R2A** converts Roman numerals to Arabic numbers (e.g., *VIII* → *8*, *CIX* → *109*, *MMXIII* → *2013*). There is no need to consider numbers beyond *MMMCMXCIX* (3999);

2. Using the previous transducers (the previous 8 transducers cannot be modified) and, if needed, other transducers, develop the following transducers:

    i. the transducer **A2R** converts Arabic numbers to Roman numbers (e.g., *8* → *VIII*, *109* → *CIX*, *2013* → *MMXIII*);

    j. the transducer **birthR2A** converts *Roman birthdates* to *Arabic birthdates* (e.g., *VIII/IX/CCCXIII* →

08/09/0313);

k. the transducer **birthA2T** converts *Arabic birthdates* to *Arabic-text format birthdates* (e.g., *08/09/0313* → *08/Sep/0313*);

l. the transducer **birthT2R** converts *Arabic-text birthdates* to *Roman birthdates* (e.g., *08/Sep/2013* → *VIII/IX/MMXIII*);

m. the transducer **birthR2L** converts *Roman birthdates* to *leap*/*not-leap* (e.g., *VIII/IX/MMXIII* → *not-leap* and *VIII/IX/MMXII* → *leap*);

3. Test the transducers **birthR2A**, **birthA2T**, **birthT2R** and **birthR2L** with the birthdate of each member of the group. The filenames of all of the files used in your tests should start with the corresponding IST student number (5 or 6 digits). For example, 12345.txt, 12345.fsm, 12345.pdf, 12345birthR2A.pdf, ...

Assume that:

- the file "syms.txt" has the symbols to be manipulated by the transducers and cannot be changed;
- the first 8 transducers cannot make conversions beyond what is required;
- you can use other transducers not mentioned in the statement;
- the days in *Arabic birthdates* and in *Arabic-text birthdates* always have 2 digits;
- the months in *Arabic birthdates* always have 2 digits;
- the years in *Arabic birthdates* and in *Arabic-text birthdates* always have 4 digits;
- there is no need to consider years beyond 3999 (MMMCMXCIX);
- when processing leap years assume the year is a 4-digit number between 1901 and 2099;
- assume that all the following inputs are valid: (i) days between 1 and 31, (ii) months between 1 and 12, (iii) years between 1 and 3999, (iv) 3-letter month names belonging to the set *{Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec}*, and (v) days and months are possible (e.g., *31/Fev* does not occur); (vi) Arabic birthdates and Arabic-text birthdates always have the correct number of digits;

Software

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

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

Submission

Submit in Fenix, project *MP1*, 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 MUST provide an estimate of each element's contribution to the work. For example: Peter: 60%, John: 40%, along with a short 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 -5 points);
2. Correct operation of the R2A transducer (4 points);
3. Correct operation of the birthR2A and birthA2T transducers (2 points each);
4. Correct operation of the leap, A2R, birthT2R and birthR2L transducers (1 point each);
5. Correct operation of the mm2mmm, d2dd, d2dddd, copy, skip and date2year transducers (0.5 points each);
6. Run.sh operating correctly (3 points);
7. Submission of the graphic versions of all transducers, as well as the examples, in their different forms, that is, after being fed as input to the transducers (1 points);
8. 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).

During the evaluation of the correct operation of any transducer, the evaluation does not take into account the origin of the errors (e.g., when testing a transducer B, every time the expected output is not obtained, an error is taken into account, even when the origin of the error is the malfunction of another transducer used to generate B). So, malfunctions on the first 8 transducers may have a huge impact on the evaluation.

"*Academic Integrity*" in Língua Natural

In this course, each student is expected to subscribe to the highest standards of academic honesty. This means that every idea that is not the student's must be explicitly accredited to the respective author. Failure to do so constitutes plagiarism.

Plagiarism includes using ideas, code, or sets of solutions from other students or individuals, or any sources other than the course texts, without crediting those sources. Students are encouraged to discuss the problems with other students and should mention this discussion when they submit their results. This mention will NOT influence the grade. Students should not, under any circumstances, show to their classmates, even temporarily, their solutions of the quizes or projects subject to evaluation. In fact, they should not even throw away drafts of the solutions without destroying them first, nor leave the developed code on shared-use computers.

Academic dishonesty also includes copying in exams. In this discipline, these should be taken without consulting any text or other classmates. Receiving or giving help during these exams is an act of academic dishonesty. Situations that could give rise to suspicions of dishonesty (opening backpacks to get paper, looking around instead of concentrating on the exam paper, etc.) should be avoided.

In this couse, academic dishonesty is considered fraud, with all the legal consequences. Any fraud will have the immediate consequence of failing all students involved (including those who enabled it to occur). Any suspicion of academic dishonesty will be reported to the higher bodies of the school for disciplinary action. This may result in failure of the subject, failure of the year, temporary or permanent suspension from IST or even from the University of Lisbon.