LÍNGUA NATURAL 2019/2020

Mini-Project Nº 2 (MP2)

Should be done:	☐ individually	☒ **in group**
Submission:	☐ theoretical class	☒ **Fenix submission**
Submission deadline:	**till 23:59, 15/Nov**	

OBJECTIVES

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

STATEMENT

Suppose you want to create a morphological analyzer, using transducers only, to couple to a text processing system. To do this, the module must convert a portuguese word into its lemma and present a morphological classification.

To simplify the project, only words that comply with the "normal" word formation rules for the nouns, adverbs and verbs of the 1st (ending in "ar") and only the following verb tenses (present, past and future) should be considered.

For non-Portuguese speakers, the necessary inflection paradigms are listed below:

a. Regular nouns (masculine singular terminated with an "o"; this form is the lemma)

Remove	Add	Tag	Example
<>	<>	+N +ms	aluno
<o>	<a>	+N +fs	aluna
<>	<s>	+N +mp	alunos
<o>	<as>	+N +fp	alunas

b. Regular Adverbs (ends with "mente"; this form is the lemma)

Remove	Add	Tag	Example
<mente>	<mente>	+ADV	inteligentemente

c. Regular Verb (infinitive ends with "ar"; this form is the lemma)

Remove	Add	Tag	Example
<ar>	<o>	+V +ip +1s	lavo
<r>	<s>	+V +ip +2s	lavas
<r>	<>	+V +ip +3s	lava
<r>	<mos>	+V +ip +1p	lavamos
<r>	<is>	+V +ip +2p	lavais
<r>	<m>	+V +ip +3p	lavam
<ar>	<ei>	+V +is +1s	lavei
<r>	<ste>	+V +is +2s	lavaste
<ar>	<ou>	+V +is +3s	lavou
<ar>	<ámos>	+V +is +1p	lavámos
<r>	<stes>	+V +is +2p	lavastes
<>	<am>	+V +is +3p	lavaram

<>	<ei>	+V +if +1s	lavarei
<>	<ás>	+V +if +2s	lavarás
<>	<á>	+V +if +3s	lavará
<>	<emos>	+V +if +1p	lavaremos
<>	<eis>	+V +if +2p	lavareis
<>	<ão>	+V +if +3p	lavarão

Where the meaning of the tags is the following:

TAG	Meaning
<+N>	noun
<+ADV>	adverb
<+V>	verb
<+ms>	masculine and singular
<+fs>	feminine and singular
<+mp>	masculine and plural
<+fp>	feminine and plural
<+ip>	indicative; present
<+is>	indicative; perfective past
<+if>	indicative; future
<+1s>	first person; singular
<+2s>	second person; singular
<+3s>	third person; singular
<+1p>	first person; plural
<+2p>	second person; plural
<+3p>	third person; plural

a. Develop the following transducers:
    1. create the transducer **lemma2noun.fst** that allows converting a noun's lemma and its tags into the corresponding form. Example: "aluno+N+fp" is converted to "alunas";
    2. create the transducer **lemma2adverb.fst** that allows converting a adjetive's lemma and its tags into the corresponding form. Example: "inteligentemente+ADV" is converted to "inteligentemente";
    3. create the transducer **lemma2verbip.fst** that allows converting a verb's lemma followed by an indicative tag into the corresponding form. Example: "lavar+V+ip+2s" is converted to "lavas";
    4. create the transducer **lemma2verbis.fst** that allows converting a verb's lemma followed by a perfective past tag into the corresponding form. Example: "lavar+V+is+2s" is converted to "lavaste";
    5. create the transducer **lemma2verbif.fst** that allows converting a verb's lemma followed by a future tag into the corresponding form. Example: "lavar+V+if+2s" is converted to "lavarás";

b. Using the previous transducers:
    6. create the transducer **lemma2verb.fst** that allows converting a verb's lemma followed by any tag into the corresponding form.
    7. create the transducer **lemma2word.fst** that allows converting any lemma followed by any tag into the corresponding form.

c. Using the previous transducers:
    8. create a transducer **word2lemma.fst** that converts any Portuguese word (ex. **alunos** (students)) into its lemma and its morphological classification (ex. **aluno+N+ms**). Sometimes a word can be classified in multiple ways (ex. **gatas** (cats) is ambiguos (ex. **gato+N+fp** and **gatar+V+ip+2s**)

d. Test the transducers **lemma2verb.fst**, **lemma2word.fst** and **word2lemma.fst** with real portuguese words but different of the examples used in this document.

Assume that:

- the file "syms.txt" has the symbols to be manipulated by the transducers and cannot be changed;
- any of the base transducers can make conversions beyond what is required;
- can use other transducers not mentioned in the statement;

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/);

An optional script is also available: *word2fst.py*, generates a transducer corresponding to a word. The *readme.txt* file explains how to use these scripts.

Submission

Submit in Fenix, in the project *MP1*, a zip file with:

- The text files used to define the transducers;
- The 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);
- All text files should use UTF8 codification;
- The final transducers should be moved to a folder named "FINALtransducers". After running the script "run.sh", this folder should contain only the following transducers: "**lemma2noun.fst**", "**lemma2adverb.fst**", "**lemma2verbip.fst**", "**lemma2verbis.fst**", "**lemma2verbif.fst**", "**lemma2verb.fst**", "**lemma2word.fst**" and "**word2lemma.fst**";
- The pdf version of the transducers should be should be kept together in a folder named "FINALpdf". After running the script "run.sh", this folder should contain only the following transducers: "**lemma2noun.pdf**", "**lemma2adverb.pdf**", "**lemma2verbip.pdf**", "**lemma2verbis.pdf**", "**lemma2verbif.pdf**", "**lemma2verb.pdf**", "**lemma2word.pdf**" and "**word2lemma.pdf**";
- Transducers with the words used to test the final transducers must have the following designations: test1.fst, test2.fst, test3.fst.The results of the question **d.** should be designated as testx_{lemma2verb, lemma2word, word2lemma}.fst;
- The pdf version of the examples: test1.pdf, test2.pdf, test3.pdf, testx_{lemma2verb, lemma2word, word2lemma}.pdf should be kept together in a folder named "FINALexamples";
- 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.

You can make several submissions, taking into account that a new submission replaces the previous one.

Attention:

- Developed transducers must have exactly the same names as above.

Evaluation criteria

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

1. Avoid unnecessary writing of transducers (1 points);
2. Correct operation of the requested transducers (1,5 points each);
3. Run.sh operating correctly (3 points);
4. 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);

5. Quality of the report [in Portuguese or in English] (1 point);
6. Report spelling and syntactic correction (1 point);

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

"Possibilities For Promoting Academic Integrity" na 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.