LÍNGUA NATURAL 2016/2017

Mini-Projecto Nº 1 — MP1

A realizar:	☐	individualmente	☒ **em grupo**
Local de trabalho:	☐	aula prática	☒ **casa**
Local de entrega:	☐	aula teórica	☒ **submissão electrónica**
Data limite entrega:	**até às 12:00 (meio dia) do dia 17/Out**		

OBJECTIVOS

Aprender a trabalhar com transdutores, usando-os para resolver um problema concreto na área de PLN.

ENUNCIADO

Os algoritmos ditos fonéticos têm como objectivo associar uma única representação (chave) a palavras que soam de modo semelhante. O *Soundex* (Odell e Russel, 1922) e o *Mataphone* (Lawrence Philips, 1990) são exemplos deste tipo de algoritmos. Por exemplo:

- JRFSK — representa Jurafsky, Jarofsky, Jarovsky e Jarovski
- NN — representa Nuno, Nunu, Nonu, Nono

Será que consegue reconhecer este texto?

AS ARMS E OS BRS ASNLDS
K D OSDNTL PR LZTN
PR MRS NNK D ANTS NVGDS
PSRM AND ALM D TPRBN

Defina UM transdutor, usando outros transdutores mais pequenos, que implemente o seguinte algoritmo fonético:

1º passo

S → *Z* se entre vogais (ex: asa → aza)

X → *Z* se precedido de "*E*" (que está no início de palavra) e seguido de uma vogal (ex: exame → ezame)

X → *S* se precedido de "*E*" (que está no início de palavra) e seguido de uma consoante (ex: extra → estra)

2º passo

CH → *X* (ex: acho → axo)
LH → *2* (ex: galho → ga2o)
NH → *3* (ex: ganho → ga3o)
RR → *4* (ex: carro → ca4o)
SS → *S* (ex: massa → masa)

3º passo

R → *4* se no início da palavra (ex: rato → 4ato)
H → ∅ (ex: hoje → oje)

Q → *K* (ex: quanto → kuanto)
Z → *S* se no fim da palavra (ex: paz → pas)
C → *S* se seguido de "*E*" ou "*I*" (ex: celofane → selofane)
K caso contrário (ex: calar → kalar)
G → *J* se seguido de "*E*" ou "*I*" (ex: gelo → jelo)
K caso contrário (ex: galo → kalo)
X → *KS* se no fim da palavra (ex: xerox → xeroks)

4º passo:

Elimina todas as vogais que não se encontram no início da palavra (ex: aleluia→ all).

Deverá ter em conta que:

- O transdutor que implementa o algoritmo fonético deverá processar uma palavra de cada vez e não uma sequência de palavras;
- O transdutor final deve ser construído usando pelo menos um transdutor por cada passo do algoritmo (4 passos);
- Dentro de cada passo é irrelevante a ordem pela qual as regras são aplicadas;
- Não serão considerados diacríticos ("´", "`", "~", "^", cedilha);
- A entrada e a saída só devem conter minúsculas;
- O ficheiro "syms.txt" contém os símbolos a manipular pelos transdutores e não pode ser alterado.

Software

Para testar a solução proposta utilize, em ambiente Linux, as ferramentas:

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

Também são disponibilizados dois scripts, de uso opcional: *compact2fst.py* (gera um transdutor a partir de outro, escrito com uma notação mais compacta) e *word2fst.py* (gerar um transdutor correpondente a uma palavra). Um ReadMe explica como usar estes scripts.

Submissão

Submeta no Fenix, no agrupamento *Mini-projeto*, um ficheiro zip (o nome do ficheiro deve ser formado por concatenação de “MP1-TAGUS” ou “MP1-ALAMEDA” (campus onde estão efetivamente a fazer a UC) com o número do grupo e com extensão “.zip”) que deve conter:

- Os ficheiros de texto usados: (i) para definir os transdutores e (ii) para teste (o último apelido de cada elemento do grupo);
- O ficheiro run.sh com TODOS os comandos usados para gerar, quer em formato FST quer em formato gráfico (PS ou PDF): (i) o transdutor final, (ii) os transdutores usados na sua construção e (iii) os exemplos de teste;
- um pequeno relatório (ficheiro txt, máximo 1 página), contendo a identificação dos elementos do grupo, a descrição das opções tomadas e comentários à solução desenvolvida (correcção e viabilidade).

ATENÇÃO: O transdutor final que implementa o conversor deverá ter o nome "transdutorFinal.fst".

Critérios de avaliação

Na avaliação serão tidos em conta os seguintes critérios no total de 4 valores):

1. Correção das soluções propostas: passos 1 e 4 (1,2 valor); passo 1, 2 e 4 (1,7 valores); todos os passos (2,0 valores);
2. Utilização de vários transdutores, pelo menos um por cada passo (0,3 valores);
3. Correto funcionamento do "run.sh" (0,5 valores);

4. Entrega dos exemplos de teste (0,3 valores);
5. Entrega das versões gráficas de todos os transdutores, incluindo os exemplos (0,3 valores);
6. Qualidade do relatório (0,3 valores);
7. Correção ortográfica e sintáctica do relatório (0,3 valores);

O não cumprimento de qualquer regra implica um desconto mínimo de 2 valores (em 4 valores).

"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.
- Make more use of the help that is available to you to master course material and to be efficient in your work:
    - Faculty and teaching assistants can talk with you during office hours or in e-mail.
    - Librarians can help you become more skilled in research.
    - Walk-in tutoring and course centers provide convenient hours to get help for several introductory courses.
    - Course bboards often include announcements and advice to aid students with particular assignments.
    - People in the class can form study groups whenever your instructor considers it appropriate so that you can practice using and explaining course concepts in a setting where peers can give you feedback.
- Recognize the options you have, other than cheating, for dealing with academic pressure:
    - Set priorities and adjust your expectations to reduce the pressure you put on yourself.
    - Talk with one of the many people on campus who may be able to offer you good suggestions: a professor, a TA, your academic advisor, your RA, or a counselor.
    - Consider participating in a workshop on time management, study skills or stress management offered by Learning Services to improve your academic success.
- If you feel tempted to cheat or plagiarize, try to identify the underlying reasons (e.g. family pressure, self-expectations, external stresses, fear of failure) and address them by talking with a friend, your parents, a counselor, your academic advisor, your TA, or someone else with whom you feel comfortable discussing the difficulties you are having.

Steps Instructors Might Take:

- Maintain clear guidelines in each course about what is permitted. You might include in course syllabi

information about ground rules and specific examples of what you consider to be cheating and plagiarism, as well as what is or isn't appropriate collaboration.

- Include in course syllabi options students can try if things get tough for them. For example, you might explicitly tell them, "Please talk with me and your TA if you are having difficulties" or "If you need some extra time to complete the assignment, the late penalty is 10% per day." Let students know that you are open to discussing extensions in exceptional circumstances and will accommodate the special needs of students with learning disabilities.
- Emphasize to students, as much as you can, that you are approachable and will try to be responsive to their concerns as individuals. Consider ways in which the course requirements could be more flexible to help students deal more effectively with academic pressure.
- Increase opportunities to discuss the course workload and get feedback on homework sets by administering early course evaluations in the first 3-5 weeks, using quality circles, and/or getting input from students via TAs throughout the semester.
- Promote more interaction in classes, especially large lectures, to increase learning. Students can avoid falling behind and will be less tempted to cheat if they have more chances to ask questions and receive help.
- Increase opportunities for one-on-one faculty-student interaction by scheduling staggered office hours, encouraging use of e-mail, and/or participating in out side-the-classroom activities. Students are much more likely to talk with you about their difficulties if they have a comfortable, ongoing relationship with you.
- Provide varied opportunities for practice and application of the course material so that the students receive frequent feedback on their performance, are aware of their standing and receive grades determined by more than just midterm and final exams. Consult the University Teaching Center if you are interested in mechanisms for increasing feedback without significantly increasing your grading load.
- Revise exams, problem sets, and paper assignments frequently. If you want to "recycle" particularly good questions, consider providing them as in-class examples, expanding a complex exam question to an out-of-class assignment, or changing elements of a question based on things like previous students' performance.
- Consider how well any graded work corresponds to the stated course objectives, how it draws on material presented in class and/or assigned reading, and whether it takes time roughly proportional to its importance in the course. Inappropriate "shortcuts" are especially tempting if students view assignments as unnecessarily long or believe exams don't test what was taught.
- Break large assignments into manageable stages to encourage effective time management, monitor and give feedback on progress, and intervene if there are potential problems. The additional structure can help students avoid last-minute acts of desperation if they fall behind.
- Clarify with your TA what responsibilities he or she has in promoting academic integrity and evaluating students' work. Discuss any specific procedures you want TAs to follow in grading and in handling a suspected violation.
- Proctor exams carefully to show your concern that all students do their work under controlled circumstances. With several monitors around the room for large classes, students also benefit by being able to ask questions more easily.
- If you suspect a student of cheating, take him or her aside and share your perceptions and concerns. Your intervention can help the student to clear up an honest misunderstanding, avoid future problems and emphasize the importance of academic integrity.