Subiecte Master anul I (Se poate folosi orice tool de pe Net, din cele discutate la laborator etc.)

Fisierul reprezentand teza se transmite la adresa de email: curs.ia.nlp@fmi.unibuc.ro

Fisierul reprezentand teza se transmite numai de la adresa institutionala a autorului.

Fisierul cu frazele (pentru exercitiile 1 si 3):
https://docs.google.com/spreadsheets/d/1hAOmYx8y3E4LdT9IMc1ILgsznejLmCLi-5R9lVw-JLU/edit?usp=sharing

Fiecare student preia fraza din dreptul numelui sau.

1. **Tabel de subsiruri bine formate (1p)**

Sa se analizeze sintactic sirul de intrare preluat din fisierul cu frazele folosind tabelul de subsiruri bine formate.

Pentru aplicarea algoritmului concepeti o gramatica CFG minimala (fara reguli in plus) si explicati semnificatia notatiilor folosite pentru neterminale (Ex.: VP = grup verbal).

Mod de rezolvare: Veti desena graful de structura a frazei. Numerotati muchiile in ordinea aparitiei lor in arbore pe parcursul analizei sintactice. Scrieti pe muchiile respective si care este regula care s-a aplicat (de ce a aparut acea muchie).

2. **Preprocesarea datelor (1p)**

Sa se explice diferenta dintre operatia de stemming si cea de lemmatizing. Care este motivul folosirii acestora?

3. **Analiza sintactica de dependenta (1p)**

Aplicati analiza sintactica de dependenta (dependency parsing), utilizand Stanford parser, sirului de intrare preluat din fisierul cu fraze.

Desenati (sub forma de graf) structura de dependenta rezultata.

4. **Algoritmul Lesk extins (2p)**

Explicati modul de utilizare a scorului de inrudire calculat de algoritmul Lesk extins pentru realizarea dezambiguizarii propriu-zise a sensului unui cuvant ambiguu (cuvantul tinta).

5. **WordNet++ (1p)**

Cum se creeaza contextul de dezambiguizare pentru o pagina de Wikipedia?

6. **Dezambiguizare cu modelul Bayes (1p)**

Comentati o modalitate de a face selectia caracteristicilor (sintactice sau semantice) in cazul dezambiguizarii locale nesupervizate care foloseste modelul Bayes.

7. **Ant Colony Algorithm (2p)**

a) Care este rolul vectorilor de miros in cadrul algoritmului? In ce situatii nu se schimba vectorul de miros al unui nod? (Enumerati minim doua situatii.)

b) Ce sunt puntile (bridges) si care este rolul lor in cadrul algoritmului?

Queries (you can use any web tool however we recommend those mentioned in the laboratory)

You must send the file containing your exam answers to this e-mail address:
curs.ia.nlp@fmi.unibuc.ro

The exam answers must be sent only from the student's institutional e-mail address.

The phrases table file (for exercises 1 and 3):
https://docs.google.com/spreadsheets/d/1hAOmYx8y3E4LdT9IMc1ILgsznejLmCLi-5R9lVw-JLU/edit?usp=sharing

Each student uses the phrase that appears next to his/her name in the table

1. **Well formed substring tables (1p)**

Perform a syntactic analysis for the given input string (yo cand find it in the phrases table) using the well formed substring tables algorithm.

Create a minimal CFG grammar (it should contain only the rules that are necessary in the analysis) and explain the meaning of the non terminal symbols (Ex: VP= verb phrase)

How to solve: Draw the parse graph for the given phrase. Add numbers to the graph connections (the numbers must indicate the order in which the oriented edges appear). For each oriented edge also write the production rule responsible for it.

2. **Text preprocessing (1p)**

Explain the difference between stemming and lemmatization. What is the role of these operations? Why do we use them?

3. **Dependency parsing (1p)**

Perform dependency parsing using the Stanford parser for the given input string (found in the phrases table).

Draw the resulting graph.

4. **Extended Lesk (2p)**

Explain how the relatedness score computed by the Extended Lesk Algorithm is effectively used for the disambiguation of a target word.

5. **WordNet++ (1p)**

How is the disambiguation context of a Wikipage created?

6. **Naive Bayes disambiguation (1p)**

Present a way of performing feature selection (of semantic or of syntactic type) for the Naive Bayes model used in local unsupervised WSD.

7. **Ant Colony Algorithm (2p)**

a) What is the purpose of the odor vectors? Find two cases in which the odor vector doesn't change for a node.

b) What are bridges and what is their role in the algorithm?