

Subiecte:

I. ANALIZA SINTACTICA

1. Analiza sintactica bazata pe constitienti. **Analiza sintactica din coltul stang** (descriere si implementare)

Bibliografie: notite curs; cartea “Introducere in procesarea limbajului natural” (Capitolul 4) sau cartea “Aspecte ale reprezentarii cunostintelor cu aplicatii in procesarea limbajului natural” (Capitolul 4)

2. Analiza sintactica bazata pe constitienti (cu memorarea rezultatelor intermediare). **Parser bottom-up cu harta** (descriere si implementare)

Bibliografie: notite curs; cartea “Introducere in procesarea limbajului natural” (Capitolul 4) sau cartea “Aspecte ale reprezentarii cunostintelor cu aplicatii in procesarea limbajului natural” (Capitolul 4)

3. **Analiza sintactica de dependenta.** Gramatici de dependenta, relatii de dependenta

Bibliografie: notite curs; cartea “Introducere in procesarea limbajului natural” (Capitolul 3; pag. 78-91)

II. ANALIZA SEMANTICA (dezambiguizare locala, dezambiguizare globala)

4. Dezambiguizarea **supervizata** a sensului cuvintelor polisemantice individuale bazata pe clasificatorul Bayesian naiv

Bibliografie: notite curs; calculul matematic pentru acest subiect (necesar formularii algoritmului) se poate gasi in cartea “Introducere in procesarea limbajului natural” (Capitolul 8, par. 8.1.1., 8.1.2., pag. 287 – 289 numai) sau cartea “Aspecte ale reprezentarii cunostintelor cu aplicatii in procesarea limbajului natural” (Capitolul 7, par. 7.1.1., 7.1.2., pag. 226-229 numai)

5. Dezambiguizarea **nesupervizata** a sensului cuvintelor polisemantice individuale bazata pe clasificatorul Bayesian naiv. Selectia caracteristicilor (sintactice, semantice)

Bibliografie: notite curs; acelasi calcul matematic de la subiectul nr. 4, la care se adauga comentariul asupra selectiei caracteristicilor; se discuta caracteristici semantice bazate pe WordNet si caracteristici sintactice bazate pe relatii de dependenta. Despre caracteristicile semantice bazate pe WordNet se poate citi in cartea “Introducere in procesarea limbajului natural” (pag. 292 - 306) sau cartea “Aspecte ale reprezentarii cunostintelor cu aplicatii in procesarea limbajului natural” (pag. 232 - 246). Despre caracteristicile sintactice bazate pe relatii de dependenta se poate citi in articolul:

Hristea, F., Colhon, M., *Feeding Syntactic Versus Semantic Knowledge to a Knowledge-lean Unsupervised Word Sense Disambiguation Algorithm with an Underlying Naïve Bayes Model*. *Fundamenta Informaticae*, Vol. 119, Nr. 1, 2012, p. 61-86.

6. Dezambiguizarea **bazata pe cunostinte** a sensului cuvintelor polisemantice individuale. Algoritmul Lesk extins prin suprapunere extinsa de glose WordNet (Banerjee si Pedersen, 2003)

Bibliografie: notite curs; cartea “Introducere in procesarea limbajului natural” (Capitolul 7, pag. 272 - 283) sau cartea “Aspecte ale reprezentarii cunostintelor cu aplicatii in procesarea limbajului natural” (Capitolul 6, pag. 212 – 223); articolul original este:

- S. Banerjee, T. Pedersen, *Extended Gloss Overlaps as a Measure of Semantic Relatedness*, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, pp. 805-810, 2003.

7. Dezambiguizare nesupervizata **globala** (a unui text) bazata pe algoritmul coloniilor de furnici (“Ant Colony Algorithm”)

Bibliografie: notite curs; articolul

- D. Schwab, J. Goulian, A. Tchechmedjiev, H. Blanchon, *Ant Colony Algorithm for the Unsupervised Word Sense Disambiguation of Texts: Comparison and Evaluation*. Proceedings of COLING 2012, Mumbai, India, pp. 2389–2404.

8. Dezambiguizarea locala **bogata in cunostinte** (bazata pe WordNet++)

Bibliografie: notite curs; articolul

- S. Paolo Ponzetto, R. Navigli, *Knowledge-rich Word Sense Disambiguation Rivaling Supervised Systems*. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 2010, pp. 1522 – 1531.