



FIȘA DISCIPLINEI

1. Date despre program

1.1 Instituția de învățământ superior	Universitatea “Alexandru Ioan Cuza” din Iași
1.2 Facultatea	Facultatea de Informatică
1.3 Departamentul	Informatică
1.4 Domeniul de studii	Informatică
1.5 Ciclu de studii	Master
1.6 Programul de studii / Calificarea	Lingvistică Computațională

2. Date despre disciplină

2.1 Denumirea disciplinei	Lanțuri de prelucrări lingvistice						
2.2 Titularul activităților de curs	Lect. Dr. Pistol Ionuț Cristian						
2.3 Titularul activităților de seminar	Lect. Dr. Pistol Ionuț Cristian						
2.4 An de studiu	2	2.5 Semestru	1	2.6 Tip de evaluare	E	2.7 Regimul disciplinei*	OB

* OB – Obligatoriu / OP – Opțional

3. Timpul total estimat (ore pe semestru și activități didactice)

3.1 Număr de ore pe săptămână	4	din care: 3.2 curs	2	3.3 seminar/laborator	2
3.4 Total ore din planul de învățământ	56	din care: 3.5 curs	28	3.6 seminar/laborator	28
Distribuția fondului de timp					ore
Studiu după manual, suport de curs, bibliografie și altele					20
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					20
Pregătire seminarii/laboratoare, teme, referate, portofolii și eseuri					20
Tutoriat [nu e cazul la facultatea noastră]					0
Examinări [evaluare pe parcurs + evaluare finala]					12
Alte activități [consultatii per student]					1
3.7 Total ore studiu individual [ST + DS + PS]					60
3.8 Total ore pe semestru [numar credite x 30 = T + E + C + TI]					129
3.9 Număr de credite					8

Obs. $T = C + S$

4. Precondiții (dacă este cazul)

4.1 De curriculum	Inteligența Artificială, Programare Java, Baze de date, Algoritmica Grafurilor, Probabilități și Statistică
4.2 De competențe	Programarea în limbaje de nivel înalt. Dezvoltarea și întreținerea aplicațiilor informatice. Proiectarea și gestiunea bazelor de date

5. Condiții (dacă este cazul)

5.1 De desfășurare a cursului	Sala de curs trebuie să dispună de video-proiector, conexiune la Internet și tablă pentru exemplificări
5.2 De desfășurare a seminarului/laboratorului	Sala de laborator trebuie să dispună de conexiune la Internet și tablă pentru exemplificări.



6. Competențe specifice acumulate [[definițiile conceptelor de mai jos se găsesc la adresa http://docis.acpart.ro/uploads/Fisiere/Metodologie%20CNCIS.pdf](http://docis.acpart.ro/uploads/Fisiere/Metodologie%20CNCIS.pdf)]

C o m p e t e n ț e p r o f e s i o n a l e	C1. Programarea în limbaje de nivel înalt. C2. Dezvoltarea și întreținerea aplicațiilor informatice. C3. Proiectarea și gestiunea bazelor de date.
C o m p e t e n ț e t r a n s v e r s a l e	CT1. Desfășurarea eficientă a activităților organizate într-un grup inter-disciplinar și dezvoltarea capacităților empatice de comunicare inter-personala, de relaționare și colaborare cu grupuri diverse

7. Obiectivele disciplinei (din grila competențelor specifice acumulate) [la fel, detalii în documentul http://docis.acpart.ro/uploads/Fisiere/Metodologie%20CNCIS.pdf](http://docis.acpart.ro/uploads/Fisiere/Metodologie%20CNCIS.pdf)

7.1 O b i e c t i v u l g e n e r a l	Utilizarea resurselor lingvistice variate în contextul unor meta-sisteme de procesare. Dezvoltarea și adaptarea unor lanțuri de procesare. Crearea de resurse adaptate utilizării în contextul unor lanțuri de procesare.
7.2 O b i e c t i v e l e s p e c i f i c e	<ul style="list-style-type: none">Lucreze în contextul aplicațiilor complexe de procesare a limbajului natural, cum ar fi în cadrul unor companii ce procesează un volum mare de date (Google, Amazon, etc.) sau în cadrul unor proiecte de cercetare de anvergură.

8. Conținut

8.1	Curs	Metode de predare	Observații (ore și referințe bibliografice)
-----	------	-------------------	--



1.	Conținut și reprezentare în Lingvistica Computațională	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [7], [8]
2.	Reprezentarea corpusurilor. Formate și instrumente de adnotare.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
3.	Reprezentarea corpusurilor. Standarde de adnotare și colecții de resurse.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
4.	Lingware. Standarde și acces.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
5.	Linguistic Processing Meta-systems	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
6.	Unstructured Information Management Architecture.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
7.	General Architecture for Text Engineering.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
8.	Integrarea și utilizarea unui meta-sistem de procesare lingvistică.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
9.	Automated Linguistic Processing Environment. Ierhia de formate și instrumente.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
10.	Automated Linguistic Processing Environment. Operații de construire a ierarhiei.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
11.	Automated Linguistic Processing Environment. Synclouds și calculul lanțurilor de procesare.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
12.	Implementarea ALPE.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
13.	Evaluarea unui meta-sistem de procesare lingvistică.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]
14.	Proiecte de cercetare.	Prezentare de slide-uri. Note de curs și tutoriale disponibile în versiune electronică.	2 ore, [1], [8], [9]

Bibliografie

Referințe principale:

- [1] Pistol.I.C. - "The Automated Processing of Natural Language", Ph.D. Thesis, UAIC, 2011
Cristea D. and Pistol I. - „Managing Language Resources and Tools Using a Hierarchy of Annotation Schemas”, in *Proceedings of the Workshop on Sustainability of Language Resources*, LREC-2008, Marakesh, 2008



- [2] Cunningham H., Maynard D., Bontcheva K. and Tablan V. - „GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications”, in *Proceedings of the 40th Anniversary Meeting of the ACL (ACL'02)*, Philadelphia, US., 2002
- [3] Ferrucci D. and Lally A. - „UIMA: an Architectural Approach to Unstructured Information Processing in the Corporate Research Environment”, in *Natural Language Engineering* 10, Number 3-4, pp. 327-348, 2004

Referințe suplimentare:

- [4] Pistol I., Trandabăț D., Iftene A., Cristea D. and Forăscu. C. - „Processing Romanian Linguistic Resources in the LT4eL Project” (in Romanian), in *Proceedings of the Workshop Linguistic Resources and Tools for Processing Romanian Language*, C. Forăscu, D. Tufis, D. Cristea (eds.), University „Al. I. Cuza” Publishing House, Iași, România, November 2006
- [5] Pistol I. C., Iftene A. - „Linguistic Processing Architecture in a GRID Environment”, in *Proceedings Scientific and Educational GRID Applications, the 5th European Conference on Intelligent Systems and Technologies (ECIT 2008)*, Publishing House "Politehniun", Iași, Editors Teodorescu H.-N. L., Craus M., pp. 61-73, ISBN 978-973-621-236-9, 10-12 July, Iași, România, 2008
- [6] Pistol I. C., Arusoae A., Vasiliu A. and Iftene A. - „A Graphical Interface for Computing and Distributing NLP Flows”, in *Proceedings of the Web Services and Processing Pipelines in HLT: Tool Evaluation, LR Production and Validation Workshop*, part of LREC 2010, available at: http://workshops.elda.org/wspp2010/sites/wspp2010/IMG/pdf/6_wspp_Pistol.pdf, 2010
- [7] Cristea D., Pistol I. C. and Forăscu C. - „ALPE as LT4eL processing chain environment”, in *Proceedings of the RANLP 2007 Workshop on Natural Language Processing and Knowledge Representation for eLearning Environments*, Paola Monachesi, Lothar Lemnitzer and Cristina Vertan (Eds.), Borovets, Bulgaria, September 26, 2007, ISBN 978-954-452-002-1, pp. 3-10, 2007
- [8] Ide N. and Romary L. - „International Standard for a Linguistic Annotation Framework”, *Journal of Natural Language Engineering*, 10:3-4, 211-225, 2004
- [9] CLARIN WP6 Group, Knowledge Sharing Infrastructure, available at <http://www.clarin.eu/external/index.php?page=publications&sub=5>, September 2010

8.2	Seminar / Laborator	Metode de predare	Observații (ore și referințe bibliografice)
1.	Recapitularea unor noțiuni: lucrul cu fișiere XML, utilizarea serviciilor Web.	Problematizare. Discuții. Schișarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [7], [8]
2.	Validarea și parsarea XML.	Problematizare. Discuții. Schișarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
3.	Adaptarea și combinarea resurselor adnotate.	Problematizare. Discuții. Schișarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
4.	Utilizarea serviciilor Web	Problematizare. Discuții. Schișarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
5.	Linguistic Processing Meta-systems	Problematizare. Discuții. Schișarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
6.	Unstructured Information Management Architecture	Problematizare. Discuții. Schișarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]



7.	General Architecture for Text Engineering.	Problematizare. Discuții. Schițarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
8.	Crearea unor plugin-uri pentru UIMA și GATE	Problematizare. Discuții. Schițarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
9.	Determinarea formatului unui document adnotat	Problematizare. Discuții. Schițarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
10.	Identitatea, transferul și combinarea documentelor adnotate	Problematizare. Discuții. Schițarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
11.	ALPE	Problematizare. Discuții. Schițarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
12.	Implementarea ALPE.	Problematizare. Discuții. Schițarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
13.	Evaluarea unui meta-sistem de procesare lingvistică.	Problematizare. Discuții. Schițarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]
14.	Proiecte de cercetare.	Problematizare. Discuții. Schițarea la tablă a soluțiilor pentru o problemă dată.	2 ore, [1], [8], [9]

Bibliografie**Referințe principale:**

- [1] Pistol I.C. - "The Automated Processing of Natural Language", Ph.D. Thesis, UAIC, 2011
- Cristea D. and Pistol I. - „Managing Language Resources and Tools Using a Hierarchy of Annotation Schemas”, in *Proceedings of the Workshop on Sustainability of Language Resources*, LREC-2008, Marakesh, 2008
- [2] Cunningham H., Maynard D., Bontcheva K. and Tablan V. - „GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications”, in *Proceedings of the 40th Anniversary Meeting of the ACL (ACL'02)*, Philadelphia, US., 2002
- [3] Ferrucci D. and Lally A. - „UIMA: an Architectural Approach to Unstructured Information Processing in the Corporate Research Environment”, in *Natural Language Engineering* 10, Number 3-4, pp. 327-348, 2004

Referințe suplimentare:

- [4] Pistol I., Trandabăț D., Iftene A., Cristea D. and Forăscu. C. - „Processing Romanian Linguistic Resources in the LT4eL Project” (in Romanian), in *Proceedings of the Workshop Linguistic Resources and Tools for Processing Romanian Language*, C. Forăscu, D. Tufis, D. Cristea (eds.), University „Al. I. Cuza” Publishing House, Iași, România, November 2006
- [5] Pistol I. C., Iftene A. - „Linguistic Processing Architecture in a GRID Environment”, in *Proceedings Scientific and Educational GRID Applications, the 5th European Conference on Intelligent Systems and Technologies (ECIT 2008)*, Publishing House "Politehniun", Iași, Editors Teodorescu H.-N. L., Craus M., pp. 61-73, ISBN 978-973-621-236-9, 10-12 July, Iași, România, 2008
- [6] Pistol I. C., Arusoiaie A., Vasiliu A. and Iftene A. - „A Graphical Interface for Computing and Distributing NLP Flows”, in *Proceedings of the Web Services and Processing Pipelines in HLT: Tool Evaluation, LR*



Production and Validation Workshop, part of LREC 2010, available at: http://workshops.elda.org/wspp2010/sites/wspp2010/IMG/pdf/6_wspp_Pistol.pdf, 2010

[7] Cristea D., Pistol I. C. and Forăscu C. - „ALPE as LT4eL processing chain environment”, in *Proceedings of the RANLP 2007 Workshop on Natural Language Processing and Knowledge Representation for eLearning Environments*, Paola Monachesi, Lothar Lemnitzer and Cristina Vertan (Eds.), Borovets, Bulgaria, September 26, 2007, ISBN 978-954-452-002-1, pp. 3-10, 2007

[8] Ide N. and Romary L. - „International Standard for a Linguistic Annotation Framework”, *Journal of Natural Language Engineering*, 10:3-4, 211-225, 2004

[9] CLARIN WP6 Group, Knowledge Sharing Infrastructure, available at <http://www.clarin.eu/external/index.php?page=publications&sub=5>, September 2010

9. Coroborarea conținutului disciplinei cu așteptările reprezentanților comunității, asociațiilor profesionale și angajatorilor reprezentativi din domeniul aferent programului

Studentii vor fi familiarizați cu utilizarea resurselor lingvistice variate în contextul unor meta-sisteme de procesare, dezvoltarea și adaptarea unor lanțuri de procesare, crearea de resurse adaptate utilizării în contextul unor lanțuri de procesare. Aceasta le permite lucrul în contextul aplicațiilor complexe de procesare a limbajului natural, cum ar fi în cadrul unor companii ce procesează un volum mare de date (Google, Amazon, etc.) sau în cadrul unor proiecte de cercetare de anvergură.

10. Evaluare

Tip activitate	10.1 Criterii de evaluare [criteriile de punctare și clasificare, inclusiv cele de promovare]	10.2 Metode de evaluare [teste scrise, proiecte, teme, prezenta (sem/lab), activitate la tabla, bonusuri pentru activități suplimentare, ...]	10.3 Pondere în nota finală (%)
10.4 Curs	Punctaj minim examen: 50% din punctajul maxim ce poate fi obținut la examen	Test scris + Bonus pentru activitatea de la curs	50%
10.5 Seminar/ Laborator	Punctaj minim laborator: 50% din punctajul maxim ce poate fi obținut la laborator (minim 50 % din punctajele celor 4 teme de laborator)	Prezentare de soluții la temele propuse	50%

10.6 Standard minim de performanță [raportate la competențele definite la punctul 7. Obiectivele disciplinei]

Studentii trebuie să fie capabili să utilizeze meta-sisteme de procesare lingvistică existente (UIMA, GATE, ALPE). Studentii trebuie să demonstreze înțelegerea mecanismului de creare a unor corpusuri și instrumente de procesare în vederea maximizării potențialului de reutilizare. Studentii vor fi capabili să adapteze resurse lingvistice existente în alte contexte decât cele pentru care au fost concepute. Proiectele propuse țin de utilizarea tehnologiilor disponibile în contexte cât mai variate. Adaptarea resurselor software și hardware presupune înțelegerea aprofundată a acestora precum și a cerințelor funcționale și consistente pentru problema curentă. Proiectele propuse au termene variabile de predare, în funcție de dificultatea lor.

Punctaj minim laborator: 50% din punctajul maxim ce poate fi obținut la laborator (minim 50 % din temele de laborator)

Punctaj minim examen: 50% din punctajul maxim ce poate fi obținut la examen

Pentru studenții care îndeplinesc criteriile de promovare nota finală se stabilește aplicând curba lui Gauss pe punctajele finale ordonate descrescător.

Studentul care participă la examen, va primi o notă, altfel va fi considerat absent. Dacă unul din criteriile de promovare nu este îndeplinit, studentul va obține o notă mai mică sau egală cu 4.



Data completării
28.09.2018

Titular de curs
Lector Dr. Pistol Ionuț Cristian

Titulari de seminar
Lector Dr. Pistol Ionuț Cristian

Data avizării în departament
28.09.2018

Director de departament
Prof. Dr. Lucanu Dorel