

FIȘĂ DE EVIDENȚĂ NR. _____
a rezultatelor activităților de cercetare-dezvoltare

TABEL NR. 1

DENUMIREA PROIECTULUI	Sistem de extragere de informații pentru limba română cu păstrarea confidențialității datelor				CATEGORIA DE PROIECT: Cercetare-dezvoltare	
CONTRACT DE FINANȚARE	225 PED/2017	DURATĂ CONTRACT	15 luni	ACRONIM PROGRAM	PN-III	
VALOAREA PROIECTULUI (INCLUDE ȘI ALTE SURSE)	627.928	VALOAREA CONTRACTULUI DE FINANȚARE (BUGET DE STAT)			474.829,85	
REZULTATELE CERCETĂRII APARTIN	ACADEMIA TEHNICA MILITARĂ „FERDINAND I” CERTSIGN SA			CONFORM: Acordului ferm de colaborare A6796 din 22.09.2017		

1) DENUMIRE REZULTAT			
2) CATEGORIA REZULTATULUI (conform art. 74, O.G. 57/2002)	Rezultat final	Rezultate intermediare	CARACTERISTICI ALE REZULTATULUI ETAPEI
2.1 documentații, studii, lucrări		×	Rezultatele sunt reprezentate de: <ul style="list-style-type: none"> • Raport privind stadiul componentelor sistemului propus (state-of-the art). • Documentație privind arhitectura sistemului propus • Site-ul web dedicat pentru prezentarea proiectului (informații actualizate la zi). • Componentele software separate din cadrul sistemului: <ul style="list-style-type: none"> - D3 - Componenta software : WebCrawler - D4 - Componenta software : Preprocesor text - D5 - Componenta software : Sistem recunoaștere entități denumite (NER) - D6 - Componenta software : Sistem extragere relații (RE) - D7 - Componenta software : ontologie + API ontologie • D8 - Software : Sistemul funcțional de extragere de informații propus în proiect (integrează componentele D3-D7).
2.2 planuri, scheme		×	
2.3 tehnologii			
2.4 procedee, metode			
2.5 produse informatice		×	
2.6 rețete, formule			
2.7 obiecte fizice/produse			
2.8 brevet invenție/alte asemenea			
3) STADIUL DE DEZVOLTARE	3.1 soluție/model conceptual		
	3.2 model experimental/ funcțional	×	

	3.3 prototip		<ul style="list-style-type: none"> Sistemul functional este reprezentat de o serie de componente legate serial (iesirea unui modul reprezinta intrarea urmatorului). Astfel, sistemul se apeleaza printr-un script in care se precizeaza folderul intrare (care va contine toate fisierele text din care vor fi extrase relatii) si folderul iesire (unde se va stoca ontologia rezultata). Sistemul apeleaza pe rand componentele necesare pentru a transforma textul din limbaj natural in noduri semantice in graful creat <p>În plus față de livrabilele asumate prin proiect au fost realizate 3 livrabile adiționale:</p> <p><u>DX1 - WordNet API pentru limba romana (componenta software)</u></p> <p>Acest API (pentru Python3) permite interogarea grafului semantic WordNet pentru limba romana. Se pot accesa literalii, cu filtrare a partii de vorbire. Se pot caută legături în graful semantic folosind de asemenea filtre pentru tipurile de relații între nodurile rețelei (ex: hypernimie, meronimie, instance-of, etc.), Menționăm că nu mai există un astfel de API pentru limba română, sau pentru a accesa datele din RoWordNet. Codul este complet documentat, există numeroase exemple de utilizare.</p> <p><u>DX2 - Corpus adnotat pentru NER (resursa lingvistica)</u></p> <p>Acest livrabil reprezinta un rezultat important în cadrul proiectului deoarece contine o resursa unică pentru Romania: un corpus cu peste 5000 de propoziții adnotate cu peste 26000 entități din 16 clase diferite. Acest corpus a stat la baza antrenării sistemului de detecție entități, și poate fi public folosit pentru a antrena sisteme ulterioare mai performante.</p> <p>Pachetul conține și ghidul de adnotare care explica metodologia de adnotare alături de exemple pentru fiecare tip de entitate.</p> <p><u>DX3 - Platforma adnotare corpus NER (componenta software adaptata)</u></p>
	3.4 instalație pilot sau echivalent		
	3.5 altele		
4) DOMENIUL DE CERCETARE	4.1 tehnologiile societății informaționale		
	4.2 energie		
	4.3 mediu		
	4.4 sănătate		
	4.5 agricultura, securitatea și siguranța alimentară		
	4.6 biotehnologii		
	4.7 materiale, procese și produse inovative		
	4.8 spațiu și securitate	✘	
	4.9 cercetări socio-economice și umaniste		
5) DOMENII DE APLICABILITATE	SECURITATE		

		Acest livrabil conține soft-ul open source BRAT (http://brat.nlplab.org) care a fost adaptat pentru a putea adnota entități denumite în cadrul proiectului. Astfel, în arhiva de pe site-ul proiectului se afla soft-ul integral, putând fi rapid instalat și utilizat pentru a extinde corpusul (o activitate de dorit în viitor, deoarece acest corpus poate fi extins ușor).	
6) CARACTERUL INOVATIV	6.1 produs nou	<input checked="" type="checkbox"/>	
	6.2 produs modernizat	<input type="checkbox"/>	
	6.3 tehnologie nouă	<input type="checkbox"/>	
	6.4 tehnologie modernizată	<input type="checkbox"/>	
	6.5 serviciu nou	<input type="checkbox"/>	
	6.6 serviciu modernizat	<input type="checkbox"/>	
	6.7 altele	<input type="checkbox"/>	
	INFORMAȚII PRIVIND PROPRIETATEA INTELECTUALĂ		
	documentațietehnico-economică	<input type="checkbox"/>	
	cerere înregistrare brevet de invenție	<input type="checkbox"/>	nr.....data.....
	brevet de invenție înregistrat (național, european, internațional)	<input type="checkbox"/>	nr.....data
	Cerere înregistrare modele și desene industriale protejate	<input type="checkbox"/>	nr.....data.....
	Modele și desene industriale protejate înregistrate (național, european, internațional)	<input type="checkbox"/>	nr.....data
	Cerere înregistrare marcă înregistrată	<input type="checkbox"/>	nr.....data.....
	Mărci înregistrate (național, european, internațional)	<input type="checkbox"/>	nr.....data
	Cerere înregistrare copyright	<input type="checkbox"/>	nr.....data.....
	înregistrare copyright (național, european, internațional)	<input type="checkbox"/>	nr.....data
	Cerere înregistrare rețele, indicații geografice, specii vegetale și animale, etc.	<input type="checkbox"/>	nr.....data.....
	înregistrare rețele, indicații geografice, specii vegetale și animale, etc. (național, european, internațional)	<input type="checkbox"/>	nr.....data

7) VALORIFICAREA REZULTATELOR CERCETĂRII								
8) DENUMIREA REZULTATULUI DE CERCETARE			SPECIFICAȚIE DE SISTEM PENTRU SRATP					
NR. CRT.	VALOAREA DE LA CARE ÎNCEPE NEGOCIEREA	PROCES VERBAL NR./DATA	MOD DE VALORIFICARE	ACTUL PRIN CARE S-A REALIZAT VALORIFICAREA	VALOAREA NEGOCIATĂ	BENEFICIAR	IMPACT	PERSOANE AUTORIZATE
0	1	2	3	4	5	6	7	8

Director proiect:
Prof.univ.dr.ing.

MIHAI TOGAN