

# NOI RESURSE LINGVISTICE COMPUTAȚIONALE ÎN MANAGEMENTUL DEZASTRELOR SOCIALE

Doctor în informatică, conferențiar universitar, cercetător științific coordonator **Mircea PETIC**  
Universitatea de Stat „Alecu Russo” din Bălți  
Institutul de Matematică și Informatică „Vladimir Andrunachievici”

## NEW COMPUTATIONAL LINGUISTIC RESOURCES IN THE MANAGEMENT OF SOCIAL DISASTERS

**Summary.** The article contains the most important aspects in creating, processing, enriching and improving the quality of the computational linguistic resources intended for the management of social disasters. The paper describes the basic features of computational linguistic resources in this field. Some working tools are presented for computational linguistic resources developed for Romanian language. Attention is drawn to essential moments in the development of qualitative computational linguistic resources.

**Keywords:** computational linguistic resources, social disasters, quality of data, crowdsourcing.

**Rezumat.** Articolul conține descrierea celor mai importante aspecte privind crearea, prelucrarea, îmbogățirea și îmbunătățirea calității resurselor lingvistice computaționale destinate managementului dezastrelor sociale. Autorul descrie particularitățile de bază ale resurselor lingvistice computaționale în acest domeniu. Sunt prezentate unele instrumente de lucru cu resursele lingvistice computaționale dezvoltate pentru limba română. Se atrage atenția asupra aspectelor esențiale de elaborare a resurselor lingvistice computaționale calitative.

**Cuvinte-cheie:** resurse lingvistice computaționale, dezastre sociale, calitatea datelor, crowdsourcing.

## INTRODUCERE

Istoria cercetărilor în domeniul procesării limbajului natural este comparabilă cu cea a calculatoarelor, dovedind în timp un interes sporit din partea oamenilor de știință, iar edificarea Societății Informaționale se constituie într-un catalizator al fenomenului dat. Astfel, interacțiunea „om-calculator” în raport cu limbajul natural este o poveste pe cât de captivantă, pe atât de relevantă. Primele încercări de procesare a limbajului natural au fost pline de optimism, după care a venit înțelegerea complexității domeniului abordat, cercetările urmând să parcurgă o cale sinuoasă.

Ultimele două decenii se remarcă printr-o ascensiune vertiginoasă a investigațiilor științifice în domeniul specificat, spectrul problemelor devenind și foarte variat. Acestea din urmă pot fi, la rândul lor, divizate în două categorii: procesarea textelor și procesarea semnalului vocal. Lucrarea de față ține de prima categorie. Realizarea tuturor acestor cercetări în bună măsură se întemeiază pe existența resurselor lingvistice computaționale de mari proporții, care, la rândul lor, stau la baza aplicațiilor ce prelucrează limbajul natural. Evident, crearea manuală a aplicațiilor ar fi un lucru extrem de laborios, de aceea este imperios necesar de a elabora metode care să contribuie la automatizarea procesului respectiv.

Pornind de la faptul că actualmente majoritatea cercetărilor și aplicațiilor sunt centrate pe limba engleză, Comisia Europeană a inițiat o serie de proiecte

în ajutorul tehnologizării celorlalte limbi europene. Or, limba română începe să se contureze ca una dintre limbile semnificative sub aspectul resurselor informaționale și a tehnologiilor ce i se aplică.

Totuși chiar și cei mai fideli susținători ai informațizării limbajului natural consideră că acest proces este încă supus hazardului. Iată de ce sunt atât de importante cercetările și elaborările care permit să înaintăm cu pași concreți în domeniul dat.

Totodată, evoluția tehnologiilor informaționale, a rețelelor și comunicațiilor, a generat un volum imens de informație textuală greu de procesat, nestructurată explicit în sens computațional, inaccesibilă sistemelor software rigide. O parte din aceste informații, și anume cele legate de dezastrele sociale (fenomene naturale distructive de origine geologică sau meteorologică, îmbolnăvirea în masă a unui număr mare de persoane sau animale, accidente cu urmări deosebit de grave asupra mediului înconjurător, acte de terorism etc.), sunt extrem de valoroase și pot fi puse în serviciul umanității pentru a preveni sau a atenua consecințele distrugătoare.

Informațiile respective cuprind foarte multe date accesibile pe rețelele globale de informare: mass-media, rețele sociale, bloguri etc. Un dezastru social este însoțit de obicei de un volum mare de date generate sub formă de știri, comentarii, dezbateri etc. Informația este destul de dificil de procesat din cauza formei sale nestructurate. Necesitatea critică de interpreta-

re, categorisire, corelare și reutilizare automatizată a acestor informații i-a făcut pe specialiștii interesați să folosească cuvintele-cheie, părțile de vorbire, relațiile semantice (de antonimie, sinonimie etc.) dintre unele cuvinte, algoritmi euristici, pentru a extrage date relevante din texte. Într-un așa mod pot fi detectate, interpretate și structurate informații de ordin social, cultural, de securitate și altele. Pentru a lua decizii corecte în timp util, sunt necesare instrumente analitice speciale. În prezent, instrumente similare există, de rând cu o gamă restrânsă de aplicații (de exemplu, monitorizarea și prognozarea incendiilor forestiere). Implementarea lor ar putea îmbunătăți în mod semnificativ modelarea dezastrelor sociale și atenuarea consecințelor acestora.

Prezenta cercetare se desfășoară în cadrul unui proiect<sup>2</sup> consacrat dezvoltării sistemelor informaționale orientate spre asigurarea securității cetățenilor în situații extreme (calamități naturale, catastrofe tehnologice etc.).

Cercetarea se referă la trei direcții principale: prelucrarea datelor textuale nestructurate în vederea stabilirii clasificatorilor pentru gruparea mesajelor emoționale; obținerea datelor calitative; prelucrarea prin mijloace crowdsourcing a datelor în vederea evaluării acestora.

### CERCETĂRI SIMILARE ÎN DOMENIU

Există numeroase lucrări care analizează subiectul avertizării dezastrelor sociale și luării deciziilor. Social media este considerată un instrument rapid de propagare a informațiilor despre catastrofele umanitare recente [1]. Acesta este motivul pentru care se utilizează frecvent în vederea monitorizării și atenuării dezastrelor [2]. Rețelele sociale oferă posibilitatea de a disemina informațiile despre pagubele produse de dezastre [3]. Un avantaj este modul de interacțiune socială pentru comunicare [4] care oferă posibilitatea postării și localizării dezastrului în rețeaua socială [5]. Mai mult decât atât, există încercări de a elabora aplicații pentru telefoane inteligente capabile să genereze soluții în cazul apariției unei situații excepționale [6].

Toate aceste soluții trebuie să ia neapărat în considerare particularitățile regionale. În [2, 7] este descris specificul problemei în România, Ucraina și Republica Moldova. Putem menționa, în special, lucrarea [2] în care se stabilește o bază a Vocabularului Controlat de limba română. Resursa dată a fost elaborată în temeiul unui șir de vocabulare controlate sub aspect profesio-

nal, precum cele ale Consiliului Internațional de Telecomunicații al Presei [8], ale autorităților specializate din România și SUA.

Un alt exemplu de vocabular profesional este prezentat în [9]. Acesta se compune din subvocabulare împărțite în 17 categorii. Fiecare categorie are mai mulți autori reprezentând autoritățile responsabile de terminologie.

Republica Moldova dispune și ea de un astfel de clasificator – cel al Serviciului Protecției Civile și Situații Excepționale [10], în care sunt descrise toate situațiile nefaste posibile în țara noastră. Pentru cercetarea respectivă, acesta servește drept bază pentru clasificarea dezastrelor sociale.

Pe de altă parte, problema completării Vocabularului Controlat este o sarcină permanentă. Cuvintele-cheie asociate dezastrelor care ar trebui incluse în acest vocabular sunt preluate din rețele sociale, bloguri, știri on-line etc. [11], ele fiind de mare importanță pentru sistemul de urgență [3].

### COLECTAREA TEXTELOR ÎN PROBLEMA-TICA DEZASTRELOR SOCIALE

Unul dintre primii pași ai cercetării a fost identificarea textelor de pe site-urile de știri, dar și de pe rețele sociale, bloguri etc. [3; 4] care conțin semnale despre dezastrul produs sau pe cale de a se declanșa undeva. În toate cazurile, algoritmi de procesare vor lua în considerare data publicării, deoarece se va lucra cu știri recente. De asemenea, vor fi excluse spoturile promoționale, zvonurile, informațiile neverificate.

Autorii proiectului au apelat la o serie de site-uri din Republica Moldova și România, colectând peste 1 000 de texte relevante în limba română, cu peste 170 de mii de cuvinte în total. Textele colectate au fost preprocesate și clasificate manual, fiind împărțite în 10 categorii de dezastre selectate din clasificatorul Serviciului Protecției Civile și Situații Excepționale: accidente feroviare, aeriene și rutiere, incendii, cutremure, uragane, poluarea cu substanțe radioactive, incidente și atacuri armate, inundații și boli. Textul care poate fi considerat ca făcând parte din două sau mai multe categorii au fost atribuite tuturor acestor categorii. Cele 10 categorii selectate acoperă circa 84 la sută din situațiile excepționale prezente în clasificator [10]. Celelalte 16 la sută din situații sunt rare în Republica Moldova, prin urmare au fost omise, ele referindu-se la pierirea în masă a animalelor sălbatice, pierirea vegetației pe un teritoriu vast, modificarea considerabilă a transparenței atmosferei etc.

Toate textele au fost lematizate și adnotate la nivel de parte de vorbire cu ajutorul POS tagger. Acest

<sup>2</sup><http://www.math.md/projects/17.80013.5007.01Ua/> - 17.80013.5007.01Ua Dezvoltarea instrumentarului pentru modelarea strategiilor de atenuare a dezastrelor sociale cauzate de catastrofe și terorism, 2017-2018.

fapt contribuie la identificarea cuvintelor care se referă la dezastrul social. Următorul pas a fost să atașăm, dacă este necesar, un segment din propoziție, pentru a evita ambiguitățile și a le clasifica. În cele din urmă, am stabilit 10 grupuri de texte raportate la cele 10 categorii de dezastru selectate, cu propriul set de markeri lexicali.

La procesarea setului respectiv de texte am aplicat și metode de dezambiguizare. Acest lucru este necesar deoarece cuvintele pot avea mai multe sensuri care apar în contexte diferite. De exemplu, cuvântul „bombă” poate avea sensul de „armă explozivă detonată în urma impactului” sau de „știre senzatională”. Un alt exemplu ar fi cuvântul „incendiar”, însemnând ceva ce cauzează (sau intenționează să provoace) incendii sau ceva care provoacă la acțiuni (de exemplu, un discurs incendiar).

## LEXICONUL MARKERILOR

După cum s-a menționat, setul de texte adnotate servește drept sursă pentru lexiconul markerilor. Cuvintele care corespund subiectului dezastrului social au fost selectate manual și adăugate la această resursă lingvistică. În cercetarea noastră [12] folosim și alte surse: site-ul Serviciului Protecției Civile și Situații Excepționale și Vocabularul Controlat de limba română, dezvoltat în [3].

Aceste trei surse constituie partea principală a lexiconului markerilor. În prezent avem un lexicon de markeri care conține peste 400 de cuvinte. Există o serie de mostre din această colecție: (a) accidenta (a răni), alarmă (alarmă), alertă, avarie, (a) bombardă, (a) deraia, (a) detona, distrugere, evacuați, (a) exploziva, (a) inunda (inundații), (a) nenoroci, (a) ucide (omorî), (a) vătăma (leza), (a) pustii (devasta) etc.

Lexiconul respectiv de markeri trebuie mereu îmbogățit cu alte cuvinte-cheie. Modalitatea obișnuită este de a spori numărul știrilor și de a selecta cuvinte noi. Pe de altă parte, pot fi utilizate mecanismele lingvistice interne pentru a spori numărul de cuvinte din lexic, începând cu setul existent. De aceea, o componentă utilă a sistemului este completarea automată a lexiconului markerilor.

### 1. Completarea automată a lexiconului

Instrumentele pe care le folosim pentru monitorizarea și analiza textului funcționează cu ajutorul cuvintelor tematice. Deoarece limba română aparține clasei de limbi flexionare, se poate produce procesul de formare a cuvintelor sau de derivare a unui număr de alternații vocale sau consoane, generând noi sensuri.

Unul dintre instrumente reprezintă aplicația Web de acces public (GeDeRo – Generatorul Derivatelor Limbii Române), care a fost dezvoltată în cadrul unui

proiect independent pentru tineri cercetători<sup>3</sup>. GeDeRo<sup>4</sup> extrage setul de derivate pentru un cuvânt anumit. Extragerea are loc în baza unui set de control al mecanismelor inteligente de gestionare interoperabilă a sistemelor informatice disponibile. Au fost utilizate următoarele resurse lingvistice computaționale relevante pentru scopurile noastre: DMLR (Dicționar Morfologic Pentru Limba Română), RRTLN<sup>5</sup> (Resurse Reutilizabile pentru Tehnologia Limbajului Natural) și eDCD (versiunea electronică a Dicționarului Cuvintelor Derivate).

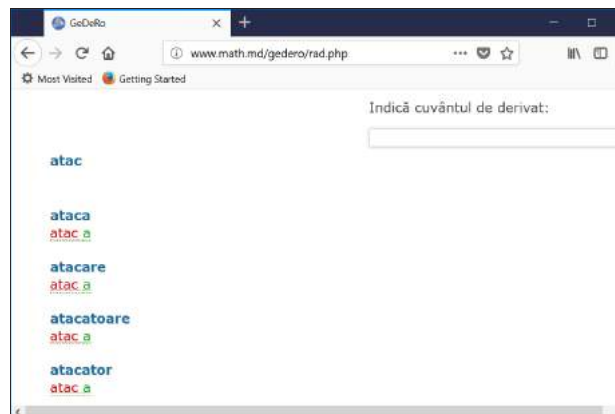


Figura 1. Exemplu pentru rădăcina cuvântului „atac”.

Interfața de pornire a aplicației Web de acces public GeDeRo oferă posibilitatea găsirii derivatelor prin prefix, rădăcină sau sufix. Având un cuvânt de intrare (de exemplu „atac”) se poate selecta un criteriu menționat mai sus și obține lista rezultatelor derivate (figura 1). Această listă este conectată la RRTLN, unde putem vedea informațiile detaliate despre cuvântul derivat (figura 2). Când facem clic pe derivate, de exemplu „atacare”, vom obține informații detaliate despre acest cuvânt (figura 1).

Formele flexionare nu își schimbă înțelesul, astfel încât acestea vor fi atribuite aceleiași clase de markeri, ca și lema cuvântului. În cazul derivării, sensul poate varia până la cel opus și este necesară o analiză suplimentară pentru a determina corespondența derivatei cu taxonomia inițială.

O altă aplicație care permite extragerea cuvintelor derivate a fost dezvoltată în cadrul proiectului pentru tineri cercetători Softcrates<sup>6</sup>. Aplicația Softcrates folo-

<sup>3</sup> Elaborarea unui suport tehnologic-informațional pentru interoperabilitatea resurselor lingvistice computaționale, 2012-2013, <http://www.math.md/projects/12.819.18.09A/>

<sup>4</sup> <http://www.math.md/gedero/> – aplicație care afișează lista de derivate în funcție de prefix, rădăcina cuvântului și sufix.

<sup>5</sup> [http://www.math.md/elrr/res\\_main.php](http://www.math.md/elrr/res_main.php)

<sup>6</sup> <http://softcrates.usarb.md/> – 15.819.02.06A Elaborarea unui sistem de prelucrare a textelor cu structură neomogenă, 2015-2016.

atacator			
atacator	substantiv (NEUTRAL)	gender	masculin
rc=122248		noun_case	Nominativ/Acuzativ
pwc=11185		number	singular
		definition	nearticulat
		proper_common_noun	comun
		syntactic_code	Substantivul la cazul <b>nominativ</b> poate avea funcția de subiect, nume predicativ, la <b>acuzativ</b> - complement direct sau indirect, complement circumstanțial de loc sau de timp.
atacator	substantiv (NEUTRAL)	gender	masculin
rc=122249		noun_case	Genitiv/Dativ
pwc=11185		number	singular
		definition	nearticulat
		proper_common_noun	comun
		syntactic_code	Substantivul la cazul <b>genitiv</b> poate avea funcția de atribut, la cazul <b>dativ</b> - de complement indirect.
atacator	substantiv (NEUTRAL)	gender	masculin
rc=122250		noun_case	Vocativ
pwc=11185		number	singular
		definition	
		proper_common_noun	comun
		syntactic_code	Substantivul la cazul <b>vocativ</b> poate avea funcția de subiect.

Figura 2. Informația detaliată pentru derivatul „atacator”.

sește serviciul Web de PoS Tagger pentru limba română<sup>7</sup> creat la Universitatea „Al. I. Cuza” din Iași.

Sistemul software în cauză, bazându-se pe mecanisme de procesare inteligentă a limbajului natural, oferă un potențial valoros de interpretare a datelor generate de utilizatorul uman într-un mod sistematic, permite folosirea acestor date în numeroase scopuri, inclusiv de analiză contextuală statistică a activității utilizatorilor software sau detectare a sensurilor textelor în limbaj natural.

Sistemul Softcrates este format din trei interfețe, fiecare corespunzând celor trei etape principale de prelucrare a materialului textual:

- captarea textului de prelucrat și numărul de rezultate scontate (figura 3);
- prezentarea celor mai des întâlnite rezultate și indicarea frecvenței acestora în textul introdus (figura 4);
- rezultatele căutării într-un motor de căutare (e.g. Google) (figura 5).

În figura 3 există o zonă textuală predestinată introducerii textului care urmează a fi analizat și prelucrat. Textul poate fi de diverse dimensiuni, dar de dorit să nu fie prea mic – nu mai puțin de două fraze. Aplicația poate prelucra informația textuală cu codificare UTF-8.

<sup>7</sup> <http://nlptools.infoiasi.ro/WebPosRo/> – Simionescu Radu, UAIC Romanian Part of Speech Tagger, 2011.

SoFTcrates - Introducerea datelor

Textul care va fi analizat

Meteorologii au prelungit codul galben de vânt până în data de 11 noiembrie. Potrivit Serviciului Hidrometeorologic de Stat, astăzi, se prevăd rafale de vânt cu viteza de până la 15 - 20 de metri pe secundă. În nordul și centrul țării sunt posibile precipitații și până la 14 grade Celsius, iar în sudul țării, temperaturile vor ajunge până la 16 grade Celsius. Luni, zece sate din cinci raioane au rămas, parțial, fără curent electric, din cauza vântului puternic. Avertizarea a fost emisă duminică de către Serviciul Hidrometeorologic de Stat.

Numarul de rezultate necesare

10

ANALIZEAZĂ

Figura 3. Interfața pentru captarea textului.

Numărul de rezultate necesare ar însemna numărul de cuvinte cu o frecvență mai mare având ca părți de vorbire substantivul, verbul, adjectivul sau adverbul. Aceasta se face pentru a extrage cuvintele cele mai sugestive. În momentul în care este apăsat butonul *Analizează* se va trece la o altă interfață, prezentată în figura 4.

Interfața prezentată în figura 4 afișează lista cuvintelor cu frecvență mai mare în textul introdus pentru prelucrare. De menționat că utilizatorul poate să selecteze cuvintele pe care le consideră mai importante și mai apropiate de conținutul textual introdus pentru analiză, nu doar din cuvintele extrase, ci și din lista de derivate asociate cuvintelor. Selectând cuvintele și apășând butonul corespunzător în motorul de căutare Google se vor găsi articole care conțin majoritatea cuvintelor indicate. În acest sens se vor găsi texte înrudite cu textul inițial.

În figura 4 este prezentată și informația cu referire la timpul consumat pentru prelucrarea textului introdus, numărul de cuvinte leme ce se repetă și numărul de cuvinte leme unice în text. La fiecare cuvânt este indicată în paranteze frecvența acestora în text. Mai mult, tot aici în interfață există posibilitatea de a descărca fișierul XML cu adnotarea morfologică a tuturor cuvintelor textului introdus pentru analiză.



Figura 4. Interfața cu informația în urma analizei textului.

În momentul în care se apasă butonul *Căutare*, va apare și rezultatul căutării în motorul de căutare Google (figura 5).

## 2. Calitatea resurselor lingvistice computaționale

Instrumentele dezvoltate sunt utile pentru crearea de resurse lingvistice computaționale care au o importanță deosebită în aplicațiile de prelucrare a limbajului natural. Construirea corpusurilor de texte mari și de bună calitate reprezintă provocarea cu care ne confruntăm astăzi [13].

Actualmente datele electronice sunt omniprezente. Din acest motiv, calitatea lor joacă un rol esențial în cele mai multe dintre tipurile de aplicații [14]. Necesitatea ca instituțiile abilitate să identifice, să evalueze și să gestioneze calitatea datelor înregistrate este din ce în ce mai vitală, pe măsură ce tehnologiile devin mai complexe, iar organizațiile se confruntă cu informații extinse și cerințe decizionale de ordin politic, social, economic și tehnic [15].

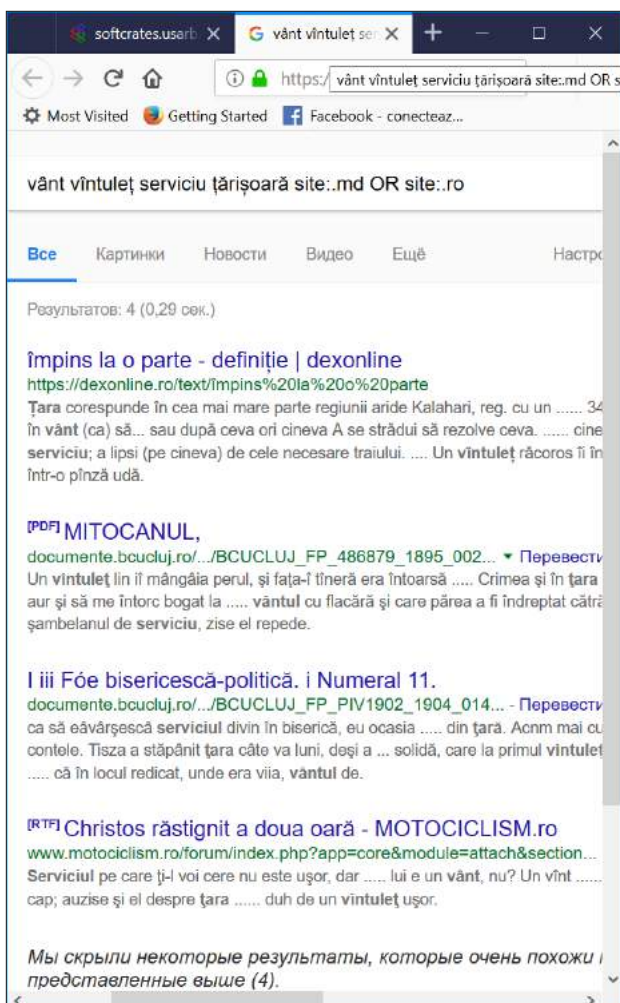


Figura 5. Interfața cu rezultatele obținute în motorul de căutare Google.

Sunt de menționat cinci factori care determină calitatea datelor.

Primul se referă la *completitudinea datelor*, adică la deficiența de date. De exemplu, am extras de pe Web că s-a produs un cutremur, dar nu am înregistrat locul producerii.

Al doilea factor este *consistența datelor* și se referă la corespunderea datelor prezentate cu cele așteptate. De exemplu, s-a semnalat producerea unui cutremur la Atena. În cazul în care în texte de limba română apare denumirea orașului scrisă în limba engleză, Athens, aceasta va limita funcționalitatea de căutare și va reduce din consistența informației selectate, dat fiind că automat vor fi excluse știrile cu transcrierea în engleză a localității respective.

Al treilea factor, *precizia datelor*, urmărește ca datele colectate să fie corecte și să corespundă realității. În privința cutremurului produs, să zicem, la aceeași Atenă, să presupunem că denumirea este scrisă corect, dar se specifică greșit epicentrul cutremurului – la 300 de km în largul Mării Egee și nu la 150.

Al patrulea factor este *validitatea datelor*, identificarea datelor nevalide însemnând adesea că există o problemă mai degrabă la nivel de proces decât de rezultat. Aceasta se determină dacă unitatea de măsură a datelor este corectă față de ce se intenționează să fie măsurat. De cele mai multe ori trebuie să fie incluse informații noi pentru ca datele să rămână valide [16].

Al cincilea factor este *actualitatea datelor* și se referă la momentul oportun de recepționare a lor pentru ca informația respectivă să poată fi utilizată în cel mai eficient mod posibil. Actualizarea permanentă a datelor este necesară pentru a preveni situația nefondată și pentru a asigura securitatea cetățenilor în situații extreme (calamități naturale, catastrofe tehnologice etc.).

Datele de calitate joasă sunt deseori asociate cu sursa de raportare precisă și cu strategii prost concepute într-o varietate de situații care apar zilnic.

### 3. Crowdsourcing-ul, o modalitate de evaluare a datelor

Indiscutabil, Web 2.0 dovedește că Web-ul îi face pe oameni disponibili drept resurse de pe urma cărora se poate profita. Web-ul contribuie la identificarea unor persoane capabile să ofere servicii prin intermediul internetului. Crowdsourcing-ul este o strategie menită să combine efortul publicului de a rezolva o problemă prin instrumente Web. Crowdsourcing-ul a fost folosit în mass-media pentru a pune în evidență faptul că angajații nu neapărat trebuie să fie experți.

Un exemplu al platformei de crowdsourcing este Amazon Mechanical Turk (<https://www.mturk.com/>). Acesta oferă forță de muncă umană la cerere, scalabilă, pentru a finaliza lucrările pe care oamenii le pot face mai bine decât computerele, precum identificarea obiectelor dintr-o fotografie sau video, dublarea datelor, transcrierea înregistrărilor audio sau cercetarea datelor în detaliu. Programul Amazon Mechanical Turk formalizează ofertele de muncă pentru mii de muncitori (cunoscuți sub numele de Turkmeni), dispuși să execute mici sarcini (denumite Sarcini de Inteligență Umană sau HIT-uri).

Deși subiecții umani pot fi implicați în furnizarea datelor sau serviciilor în mai multe forme, în cadrul proiectului ne-am limitat la analiza calității datelor privind dezastrele sociale declanșate [17].

## CONCLUZII

Cu toate că există numeroase lucrări științifice în domeniul avertizării și luării deciziilor privind dezastrele sociale, domeniul reclamă studii aprofundate și analiza unui volum extins de resurse lingvistice computaționale.

Compilarea resurselor lingvistice computaționale pe scară largă necesită o abordare atentă și subtilă, ele fiind valoroase doar atunci când calitatea resurselor respective este foarte bună.

Verificarea calității resurselor lingvistice computaționale reprezintă o zonă care rămâne mai aproape de activitatea umană decât de prelucrarea computerizată. Acesta este motivul pentru care o abordare tranșantă contribuie la obținerea datelor calitative prin intermediul resurselor lingvistice computaționale pentru o utilizare ulterioară în aplicațiile de prelucrare a limbajului natural.

## BIBLIOGRAFIE

1. Hodas N. O., Steeg G. V., Chikkagoudar S., Harrison J., Bell E., Corley C. D. Disentangling the Lexicons of Disaster Response in Twitter. In: WWW 2015 Companion, Florence, Italy, May 1822, 2015, p. 1201-1204.
2. Teodorescu H. N. Using analytics and social media for monitoring and mitigation of social disasters. In: Procedia Engineering, vol. 107C, 2015, p. 325-334.
3. Gowri A. S., Kavitha R. Tweet Alert: Effective Utilization of Social Networks for Emergency Alert and Disaster Management System. In: International Research Journal of Engineering and Technology, vol. 2, no. 8, 2015, p. 1065-1070.
4. Simionescu R. Hybrid POS Tagger. In: Proceedings of Language Resources and Tools with Industrial Applications Workshop (EuroLan 2011 Summer School). Cluj-Napoca, Romania, 2011, p. 21-28.
5. Rana R., Kristiansson I., Hallberg I. and Synnes K. An Architecture for Mobile Social Networking Applications. In: First International Conference on Computational Intelligence Communication Systems and Networks, CICSYN '09, 2009, p. 241-246.
6. Brownlee and Liang Y. Mobile Ad Hoc Networks: An Evaluation of Smart phone Technologies, Kingston (ONTARIO). Canada: Royal Military College of Canada, 2011, 40 p.
7. Pankratova N. D., Bidyuk P. I., Selin Y. M., Savchenko I. O., Malafeeva L. Y., Makukha M. P. and Savastiyarov V. V. Foresight and Forecast for Prevention, Mitigation and Recovering after Social, Technical and Environmental Disasters. In: Improving Disaster Resilience and Mitigation – IT Means and Tools, Springer, 2014, p. 119-134.
8. Hjelm J. Why IPTV: Interactivity, Technologies, Services, John Wiley & Sons, 2008, 370 p.
9. U.S. Department of Health & Human Services. Disaster Information Management Research Center. Disaster Glossaries, [Online]. Disponibil: <https://disaster.nlm.nih.gov/dimrc/glossaries.html>
10. Inspectoratul General pentru Situații de Urgență, Clasificator, [Online]. Available: <http://www.dse.md/ro/clasificator>
11. Pankratova N. D. and Dozirtsiv V.O. Application of

methods for text analysis of the emotional tone to identify social disasters. In: System analysis and information technology: 18-th International conference SAIT 2016, Kyiv, Ukraine, May 30 – June 2, 2016, p. 38.

12. Petic M., Cojocaru S. and Gisca V. Exploring list of markers in automatic unstructured text data processing. In: Proceedings of the 11th International Conference Linguistic Resources and Tools for Processing the Romanian Language, Iași, Romania, 26-27 November 2015, p. 125-136.

13. Petic M., Cozlov V. Determining emotional classifiers for social disasters text clustering. In: Conference on Mathematical Foundations of Informatics. Proceedings MFOI2017, November 9-11, 2017, Chisinau, Moldova, p. 146-149

14. Batini C. et al. Methodologies for Data Quality Assessment and Improvement. In: Journal ACM Computing Surveys (CSUR) Surveys Homepage archive Volume 41 Issue 3, July 2009, Article No. 16.

15. Palmer B. Ch. An approach for managing data quality. In: Master Thesis, Faculty of Information Sciences and Engineering, University of Canberra, Australia, November 2011. 155 p.

16. 5 Factors of High Quality Data & How They Affect Business Decisions, <https://nektardata.com/5-factors-of-high-quality-data/> (vizitat 18.09.2018).

17. Wang A., Hoang C., Kan M. Perspectives on Crowdsourcing Annotations for Natural Language Processing, Technical report, The National University of Singapore, July, 2010, 24 p.



Ada Zevin. *Casă în Bairac*, 1976–1983, u/p.