

Analiza trăsăturilor psihologice a autorului mesajelor postate online

Irina IURCIUC Victoria BOBICEV

Technical University of Moldova

irina.iurciuc.eni@gmail.com, victoria.bobicev@mail.utm.md

Rezumat — Studiile în domeniul psihologiei au demonstrat că particularitățile psihologice afectează direct modul în care omul comunică, scrie, alege unele sau alte cuvinte pentru a se exprima. Lucrarea dată este dedicată dezvoltării unui sistem ce analizează mesajele postate online și crează portretul psihologic al autorului. Astfel de aplicație este o inițiere în domeniul dat și un suport pentru cercetările ulterioare din viitorcare ar putea fi de mare ajutor pentru o gamă largă de instituții și servicii: de la procesul de interviuare și angajare, controlul și managementul resurselor umane într-o întreprindere, până la utilizarea softului de către agențiile guvernamentale și de securitate.

Index Terms — analiza personalității, modelul Big Five, analiza textului, caracteristici lexicale, aplicație online.

I. INTRODUCERE

Studiile în domeniul psihologiei au demonstrat că particularitățile psihologice afectează direct modul în care omul comunică, scrie, alege unele sau alte cuvinte pentru a se exprima. Astfel, dezvoltarea unui sistem ce analizează mesajele postate online și crează portretul psihologic al autorului ar putea fi de mare ajutor pentru o gamă largă de instituții și servicii: de la procesul de interviuare și angajare, controlul și managementul resurselor umane într-o întreprindere, până la utilizarea softului de către agențiile guvernamentale și de securitate pentru prezicerea și prevenirea infracțiunilor, suicidurilor și altor divergențe în societate.

Personalitatea este atributul uman cel mai complex și dificil de înțeles. Ea descrie, de fapt, unicitatea unei persoane. În ultimii ani, interesul comunității științifice în recunoașterea personalității a crescut imens. Aspectele umane psihologice au atras interesul cercetătorilor din diverse domenii, ca lingvisticii computaționale, psihologiei, inteligenței artificiale, prelucrării limbajului natural, interacțiunii dintre om și mașină, analiza comportamentală și învățarea automată.

II. MODELUL CELOR CINCI FACTORI: BIG FIVE

Modelul celor cinci factori ai personalității reprezintă o organizare ierarhizată a trăsăturilor de personalitate sintetizate pe următoarele dimensiuni [2]:

- 1.Extraversie,
- 2.Agreabilitate,
- 3.Conștiințiozitate,
- 4.Neuroticism și
- 5.Deschidere către experiență.

Încă din anii 1980 cercetătorii au ajuns la concluzia că factorii Big Five sintetizează dimensiunile fundamentale ale personalității. Păreră unanimă a teoreticienilor în legătură cu modelul este aceea că cei cinci factori se regăsesc într-o formă sau alta în toate instrumentele care evaluează personalitatea. Așadar, acest model conceptual este unul dintre cele mai solide din punct de vedere

științific, din domeniul psihologiei. Descrierea celor cinci dimensiuni este prezentată în tabelul 1.

Modelul a rezultat în urma a două modalități de analiză: cea lexicală (prin intermediul limbajului uzual au fost identificate cele mai comune trăsături de personalitate, identificate în mai multe medii culturale) și analiza chestionarelor.

Abordarea aceasta, numită și psiholexicală, enunță că diferențele deosebit de semnificative și importante în stabilirea și menținerea relațiilor interumane sunt codificate în limbaj și se exprimă prin acei termeni lingvistici care au semnificație pentru descrierile comportamentului uman [1].

TABELUL I. CELE CINCI DIMENSIUNI ALE PERSONALITĂȚII

Scor scăzut	Dimensiunea	Scor ridicat
coborât la pământ, necreativ, convențional, necurios	Neuroticism	imaginative, creative, curios, original
neglijent, lenos, nepunctual, dezordonat	Deschiderea spre experiență	conștient, organizat, sîrguincios, punctual
singuratic, tăcut, pasiv, rezervat	Extraversie	de companie, vorbăreț, active, afectuos
suspicios, critic, iritabil, nemilos	Agreabilitate	credul, îngăduitor, inimos, bun de fire
calm, temperat, confortabil, neemoționat	Conștiințiozitate	îngrijorat, temperamental, conștient, emoțional

III. ANALIZA AUTOMATĂ A PERSONALITĂȚII

Numărul rețelelor sociale a crescut dramatic în ultimul deceniu. Acestea au devenit utilizate pe scară largă și în diverse domenii atât pentru comunicare și interacțiunile sociale, cât și pentru formarea opiniilor publice, atragerea clienților, distribuția informațională. Activitățile utilizatorilor pe astfel de site-uri oferă posibilitatea de a observa comportamentul individual al utilizatorilor, a experiențelor, a opiniilor și a intereselor. Informația este liberă pentru vizualizare și demonstrează modul în care o persoană interacționează social cu alți oameni.

Există un număr de aplicații care oferă serviciul de recunoaștere a personalității sau mai concret a trăsăturilor psihologice care definesc o personalitate. Aceste analize au

loc în baza informației survenite pe rețelele sociale, forumuri, în cadrul întreprinderilor și abordează diverse aspecte precum detectarea înșelăciunilor, analiza sentimentelor, prevenirea acțiunilor cu caracter negativ etc.

O gamă largă de trăsături lingvistice asociate cu fiecare dintre cele cinci dimensiuni ale personalităților au fost identificate în [5]. Autorii cercetării au folosit resursele limbii engleze Lingvistic Inquiry și Word Count (LIWC) pentru a număra categorii de cuvinte în eseurile scrise de studenți a căror personalitate a fost evaluată folosind un chestionar.

În [6] a fost prezentat un studiu personalităților în baza textelor și au fost descrise pașii necesare pentru astfel de cercetare:

1. Colectarea textelor persoanelor pentru cercetare;
2. Estimarea caracteristicilor personalităților pentru fiecare participant;
3. Extrageți caracteristicile relevante din texte;
4. Construirea modelului statistic ale dimensiunilor personalităților bazate pe caracteristici lingvistice;
5. Testarea modelului învățat pe caracteristici lingvistice.

În calitate de caracteristici extrase din texte au fost folosite cuvintele din diferite domenii, cum ar fi, mâncarea, spațiu, sentimente, etc., semne de punctuație, cuvinte funcționale: pronume, conjuncții, prepoziții. Rezultatul testării a demonstrat că elementele textului pot fi utilizate pentru evaluarea trăsăturilor personalității autorului textului.

Extraversiunea și legătura ei cu utilizarea limbii a fost studiată în [4]; concluziile fiind că aceasta este corelată în mod semnificativ cu folosirea multor expresii deictice, cum ar fi pronumele, verbele, adverbele și interjecții.

Michal Kosinski a creat myPersonality Project Website (<http://mypersonality.org/wiki/doku.php>) care prezintă o Aplicație Facebook ce permite utilizatorilor să efectueze teste psihometrice reale [8]. În prezent, baza de date a aplicației conține mai mult de 6.000.000 de rezultate, împreună cu peste 4.000.000 de profiluri Facebook individuale.

Fezabilitatea modelării personalității utilizatorilor pe baza unui set propus de caracteristici extrase din datele de pe Facebook au fost explorate în [7]. În calitate de resurse cercetate au fost utilizate datele despre 250 persoane, în total 10000 statusuri și postări obținute din setul descris în [8]. Din datele acestea au fost extrase 725 caracteristici de două tipuri. Unele descriu persoana în general, ca de exemplu, datele generale de pe profilul facebook. Alt tip de caracteristici avea caracter lingvistic și a fost extras direct din textele postate de persoanele respective. Cercetarea a demonstrat creșterea strânsă între cuvintele utilizate și trăsăturile personale. Cele mai importante cuvinte au fost adjectivele și cuvintele din dicționarele afective General Inquirer (www.wjh.harvard.edu/~inquirer/Home.html) și Afinn (http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010).

Senticnet [10], o resursă semantico-lexicală a fost utilizată în [9] cu același scop: detectarea personalității autorului pe baza textului. Cinci modele statistice au fost

create pentru cele cinci trăsături de personalitate, Autorii au concluzionat că utilizarea cunoștințelor de bun simț cu informații afective și sentimente duce la rezultate sporite în detectarea personalității.

IBM Watson Personalitate Insights Service [3] încearcă să utilizeze analiza lingvistică pentru a determina dintr-o bucată de text portretul psihologic al autorului său. Textul este analizat în baza anumitor cuvinte marcate, iar rezultatele sunt sortate după metrica psihologică Big Five. Serviciul are API, documentația la dispoziția publicului, de asemenea o secțiune pe GitHub. Dar, există un mic dezavantaj: textul trebuie să fie în limba engleză sau spaniolă.

IV. CREAREA APLICAȚIEI PENTRU ANALIZA PERSONALITĂȚII

Luând în considerație actualitatea temei investigate și utilitatea unui instrument ce ar evalua particularitățile caracteriale și temperamentale ale unei persoane, sarcina și scopul fundamental al acestei lucrări este elaborarea unei aplicații destinate pentru analiza trăsăturilor psihologice ale autorului în baza mesajelor postate online. Aplicația presupune capacitatea de a prelua o colecție de texte și de a le analiza din punct de vedere al vocabularului utilizat și a manierei de exprimare.

Inițial au fost stabilite cerințele către aplicația:

- Textele pentru analiză pot fi introduse atât manual, cât și sub formă de fișier text.
- Informația privind apartenența la una din cele cinci dimensiuni psihologice ale modelului "Big Five" trebuie să fie cât mai clară și cât mai corect formulată;
- Portretul psihologic stabilit de aplicație trebuie să fie cât mai aproape de adevăratul portret psihologic al persoanei care o utilizează;
- Interfața aplicației trebuie să fie prietenoasă și pe cât posibil de laconică, pentru a evita dificultăți în utilizare.

Întrucât primul pas în crearea aplicației ce analizează texte este colectarea resurselor, și anume textelor propriu zise, inițial au fost colectate postările de pe Facebook. Figura 1 prezintă un exemplu de postări a unei persoane.

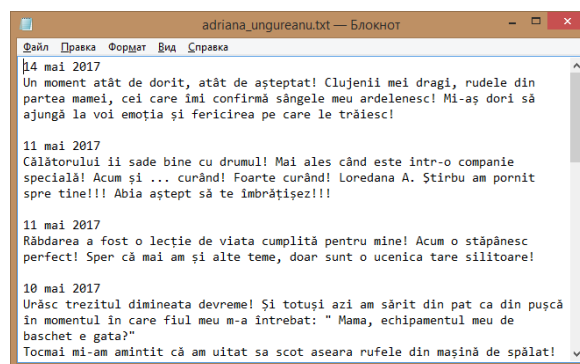


FIGURA 1. EXEMPLU DE POSTARI A UNEI PERSOANE PE FACEBOOK

Textele colectate sunt procesate conform modelului Big Five a celor cinci dimensiuni psihologice ale personalității. Cu scopul de a oferi o metodă eficientă și efectivă pentru

studierea diferitelor componente emoționale, cognitive și structurale prezente în limba verbală și scrisă ale omului, a fost creată inițial o aplicație de analiză a textului numită anchetă lingvistică sau LIWC (Linguistic Inquiry and Word Count).

În proiectarea și dezvoltarea LIWC, au fost generate mai întâi seturi de cuvinte pentru fiecare categorie. În urma creării unor liste preliminare de cuvinte, s-au organizat sesiuni de brainstorming între 3-6 "judecători", care au creat cuvinte relevante pentru diferitele categorii. Creatorii LIWC de asemenea au analizat ponderile de utilizare ale anumitor categorii de cuvinte pentru anumite categorii de personalitate.

Astfel, au fost formate tabele de ponderi pentru diferite categorii lexicale ce reflectă utilizarea lor de persoane cu caracteristicile respective. Tabelul II prezintă un exemplu de categorii lexicale cu ponderi asociate pentru fiecare dimensiune din cele Big Five: 1 - Extraversie, 2 - Agreabilitate, 3 - Conștiinciozitate, 4 - Neuroticism și 5 - Deschidere către experiență.

TABEL II. FACTORII LIWC ȘI CORELAȚIILE LOR CU DIMENSIUNILE MODELULUI BIG FIVE

LIWC factor	1	2	3	4	5
Cuvinte la persoana întâi singular	.04	.07	.01	.13	-.13
Articole	-.09	-.15	-.04	.09	.13
Cuvinte cu mai mult de 6 litere	-.04	-.03	.06	-.03	.16
Verbe la timpul prezente	.01	.04	.00	.06	-.15
Cuvinte discrepantă	-.03	-.02	-.07	.05	-.01
Cuvinte exclusive	-.08	-.06	-.08	.00	.10
Cuvinte tentativă	-.14	-.02	-.06	.06	.11
Negații	-.12	-.04	-.15	.05	.00
Cuvinte inclusive	.07	.03	.06	-.01	.01
Verbe la timpul trecut	.04	.06	-.06	.03	-.03
Cuvinte sociale	.12	.00	.02	-.01	.02
Cuvinte ce exprimă emoții pozitive	.15	.07	.07	-.13	-.06
Cuvinte ce țin de înțelegere	-.02	.05	-.01	-.06	.07
Cuvinte cauzale	-.08	.00	-.07	.03	-.08
Cuvinte ce exprimă emoții negative	-.08	-.07	-.15	.16	.05

Din tabelul dat putem observa, de exemplu, că persoanele predispuse spre neuroticism folosesc multe cuvinte la persoana întâi singular, multe cuvinte negative, puține cuvinte exclusive. Totodată neuroticii utilizează

foarte puține articole, cuvinte cu mai mult de 6 litere, cu atât mai mult cuvinte care exprimă emoții pozitive.

Luând în considerație faptul că pentru limba română nu există astfel de sursă unde toate cuvintele să fie clasificate în categoriile menționate în tabelul II, am luat în considerație câteva din aceste categorii de cuvinte ca criterii de stabilire a apartenenței la una dintre cele cinci categorii ale modelului Big Five: neuroticism, extraversiune, deschidere spre experiență, agreabilitate și conștiinciozitate. În tabelul III sunt evidențiate categoriile din tabelul precedent care au fost folosite în aplicația creată.

Semnul plus înseamnă că în textul analizat este cel puțin un cuvânt din categoria respectivă sau, mai corect spus, cuvintele din categoria dată trebuie să constituie un procentaj mai mare ca zero din numărul total de cuvinte analizate.

TABEL III. FACTORII LIWC ȘI CORELAȚIILE LOR CU DIMENSIUNILE MODELULUI BIG FIVE UTILIZATE ÎN APLICAȚIA CREATĂ

LIWC factor	1	2	3	4	5
Cuvinte la persoana I singular	+	+	+	+	-
Articole	-	-	-	-	+
Cuvinte cu mai mult de 6 litere	-	-	+	-	+
Cuvinte tentativă	+	+	+	+	-
Verbe la timpul prezent	-	-	-	+	+
Verbe la timpul trecut	+	+	-	+	-
Cuvinte care exprimă emoții pozitive	+	+	+	-	-
Cuvinte care exprimă emoții negative	-	-	-	+	+

Aplicația creată are la bază următorul algoritm: dacă o categorie de cuvinte este prezentă în text, atunci punctajul pentru una dintre cele 5 dimensiuni se incrementează.

Dacă în text procentajul de cuvinte la persoana I singular este mai mare ca zero, punctajul pentru Extraversie se incrementează. Același lucru se respectă pentru procentajul cuvintelor tentativă, procentajul verbelor la timpul prezent, procentajul verbelor la timpul trecut și procentajul cuvintelor care exprimă emoții negative. Punctajul maxim pentru extraversiune, de exemplu, poate fi 4, pentru neuroticism - 5.

Însă verificările menționate nu sunt suficiente pentru a stabili apartenența la una dintre cele cinci dimensiuni, de exemplu, pronume la persoana I singular incrementează patru din cinci dimensiuni în același timp. Astfel, conform tabelului LIWC, am mai stabilit un șir de reguli și condiții, care, fiind îndeplinite, de asemenea incrementează punctajul pentru o dimensiune anumită.

Un exemplu de condiție este: punctajul pentru Agreabilitate se mărește doar dacă numărul cuvintelor la persoana I singular este mai mare ca numărul de verbe la timpul prezent.

TABEL IV. EXEMPLE DE CUVINTE DIN LEXICOANELE CREATE

Categoria de cuvinte	Exemple
Cuvinte la persoana I singular	eu, mine, mie
Articole	un, o, unul
Cuvinte tentativă	aș face, pot încerca
Verbe la timpul prezent	merge, citesc
Verbe la timpul trecut	vedeam, a spus
Cuvinte care exprimă emoții pozitive	vesel, bucura, fericire
Cuvinte care exprimă emoții negative	trist, urî, durere

Cele opt categorii de cuvinte prezentate în tabelul III servesc drept criterii de analiză a textului și sunt repartizate în lexicoane. și stocate în fișiere text cu extensia “.csv”. Aplicația analizează textul cuvânt cu cuvânt și numără câte din ele se conțin în lexiconul respectivei categorii. De exemplu, în procesul de numărare a procentajului de articole din numărul total de cuvinte, aplicația parcurge textul și lexiconul cu numele “lexArticole.csv”. În momentul în care întâlnește un cuvânt care se conține în text și în lexiconul respectiv, numărul de articole crește cu o unitate. Același procedeu are loc și cu restul lexicoanelor. Exemple de cuvinte din categoriile date sunt prezentate în tabelul IV.



FIGURA II. INTERFAȚA APLICAȚIEI

Astfel, utilizând lexicoanele respective și regulile

menționate mai sus este evaluată personalitatea autorului textului prezentat pentru analiză.

V. CONCLUZII

Crearea unei aplicații de analiză a portretului psihologic în baza textelor în limba română a pus bază în inițierea în domeniu, prezintă și asigură suport pentru cercetările ulterioare din viitor.

REFERENCES

- [1] Macarie A. Constantin T., Orzan A., Constantin L. Fodorea A., Modelul Big Five al personalității; abordări teoretice și modelarea empirică a unui chestionar standardizat., în Revista „Psihologia Resurselor Umane”, vol 6, nr. 2/ 2008 ,(pp. 62 – 74)
- [2] Nina Sofian, Despre studiul personalității: Istoricul modelului Big Five, sau ce se află în spatele celor mai bune teste de personalitate, 2014. Accesat 13.09.2017.
- [3] IBM Watson Personality Insights <https://www.ibm.com/watson/developercloud/personality-insights.html>
- [4] Francois Mairesse, Marilyn A. Walker , Matthias R. Mehl, Roger K. Moore, Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text, Journal of Artificial Intelligence Research, Volume 30 Issue 1, 2007, Pages 457-500.
- [5] Pennebaker, J. W., & King, L. A. Linguistic styles: language use as an individual difference. Journal of Personality and Social Psychology, 77(1999):1296–1312.
- [6] Mairesse, F, Walker, MA. Words Mark the Nerds: Computational Models of Personality Recognition through Language. 28th Annual Conference of the Cognitive Science Society, 2006.
- [7] D. Markovikj, S. Gievaska, M. Kosinski, D. Stillwell. Mining Facebook Data for Predictive Personality Modeling, Proceedings of AAAI International Conference on Weblogs and Social Media (ICWSM), 2013.
- [8] Celli, F., Pianesi, F., Stillwell, D. S., and Kosinski, M. Workshop on Computational Personality Recognition (Shared Task). The Seventh International AAAI Conference on Weblogs and Social Media. USA, 2013.
- [9] Soujanya Poria, Alexandar Gelbukh, Basant Agarwal, Erik Cambria, Newton Howard. Common Sense Knowledge Based Personality Recognition from Text. Advances in Soft Computing and Its Applications, 2013, pp 484-496.
- [10] Cambria, E., Havasi, C. and Hussain, A. “SenticNet 2: A semantic and affective resource for opinion mining and sentiment analysis.” Proceedings of FLAIRS, Marco Island, pp.202–207, 2012.