



(12) CERERE DE BREVET DE INVENȚIE

(21) Nr. cerere: a 2021 00184

(22) Data de depozit: 19/04/2021

(41) Data publicării cererii:  
30/12/2021 BOPI nr. 12/2021

(71) Solicitant:  
• UNIVERSITATEA POLITEHNICA DIN  
BUCUREȘTI, SPLAIUL INDEPENDENȚEI  
NR.313, SECTOR 6, BUCUREȘTI, B, RO

(72) Inventatori:  
• RUȘEȚI ȘTEFAN, STR.DOAMNA GHICA,  
NR.32B, BL.T3, AP.608, SECTOR 2,  
BUCUREȘTI, RO;

• NEAGU LAURENȚIU-MARIAN,  
STR.VALEA GLĂMBOCULUI, NR.61,  
COMUNA BASCOV, PITEȘTI, AG, RO;  
• TOMA IRINA, STR.ION HELIADE  
RĂDULESCU NR.64, CĂLĂRAȘI, CL, RO;  
• DASCALU MIHAI,  
STR.STOICA LUDESCU, NR.61, ET.1, AP.7,  
SECTOR 1, BUCUREȘTI, B, RO

(54) METODĂ DE ÎNVĂȚARE AUTOMATĂ DE REPREZENTĂRI  
VECTORIALE ÎN GRAFURI DE CUNOȘTINȚE PORNIND  
DE LA MODELE DE LIMBĂ

(57) Rezumat:

Invenția se referă la o metodă de învățare a reprezentărilor vectoriale ale unor entități dintr-un graf de cunoștințe care îmbină descrierea textuală asociată entității cu reprezentări vectoriale derivate din graful de cunoștințe. Metoda conform invenției cuprinde două faze, o fază de antrenare și o fază de analiză a informațiilor folosind un nou spațiu de reprezentare, reprezentarea vectorială a unei entități fiind construită pe baza relațiilor existente în graful de cunoștințe, precum și pe o posibilă descriere textuală a entității, procesată folosind modele delimbă pre-antrenate, cum ar fi BERT, sau alte modele semantice de reprezentare vectorială a textelor. În faza de antrenare, modelul calculează, folosind o rețea neurală, reprezentarea unui nod prin combinarea descrierii textuale vectorizate cu reprezentarea nodului modificată la antrenare, urmată de proiecția entităților dintr-un triplet de forma <nod sursă, relație, nod destinație> în spațiul de reprezentare al relației dintre noduri, calculul distanței dintre nodul sursă și relație, raportat la nodul de destinație, respectiv actualizarea ponderilor pe baza distanțelor calculate prin tehnica de "backpropagation". În faza de analiză sunt identificate entități similare și noi relații, respectiv sunt vizualizate reprezentările vectoriale ale entităților.

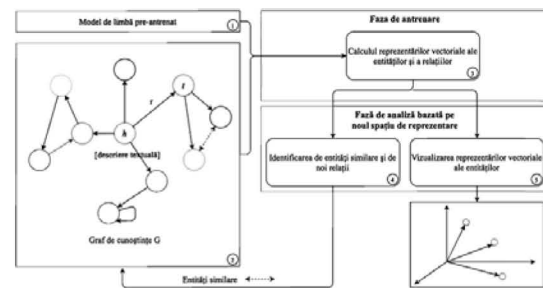


Fig. 1

Revendicări: 5  
Figuri: 3

Cu începere de la data publicării cererii de brevet, cererea asigură, în mod provizoriu, solicitantului, protecția conferită potrivit dispozițiilor art.32 din Legea nr.64/1991, cu excepția cazurilor în care cererea de brevet de invenție a fost respinsă, retrasă sau considerată ca fiind retrasă. Întinderea protecției conferite de cererea de brevet de invenție este determinată de revendicările conținute în cererea publicată în conformitate cu art.23 alin.(1) - (3).



OFICIUL DE STAT PENTRU INVENȚII ȘI MĂRCI
Cerere de brevet de invenție
Nr. <u>a 221 00189</u>
Data depozit <u>19-04-2021</u>

## Metodă de Învățare Automată de Reprezentări Vectoriale în Grafuri de Cunoștințe pornind de la Modele de Limbă

### Descriere

#### Prezentarea domeniului de aplicare

Invenția se referă la o metodă de învățare automată a unei reprezentări vectoriale pentru fiecare entitate dintr-un graf de cunoștințe, luând în considerare atât descrierile în limbaj natural ale nodurilor, cât și relațiile dintre noduri. Reprezentările vectoriale codifică proprietăți latente ale grafului, iar obiectivul final este ca entități similare să fie descrise sub forma unor vectori similari. Domeniul de aplicare al metodei este vast, aceasta contribuind asupra tehnicilor de prelucrare a limbajului natural, prin îmbunătățirea aplicațiilor ce includ analiză semantică, căutări semantice, sisteme de recomandări sau agenți conversaționali / chatbots. Metoda poate fi aplicată pentru a determina relații între entitățile similare și acoperă orice domeniu în cadrul căruia cunoștințele pot fi modelate sub forma unui graf de cunoștințe, ontologii, iar descrieri textuale sunt disponibile la nivelul entităților; fără a ne limita, menționăm: cercetarea documentară (rapoarte științifice), jurnalistică, istorie, medicină, sau beletristică. Metoda permite totodată determinarea entităților similare de același tip (spre exemplu, orașe, scriitori, evenimente istorice) pornind de la descrierile textuale ale entităților și un graf de cunoștințe care modelează relații între entități, generând asocieri de tipul “entitatea  $x$  este similară cu entitatea  $y$ ”.



## Stadiul tehnicii

Brevetul CN107665252B (<https://patents.google.com/patent/CN107665252B/en>) prezintă o metodă de generare a grafurilor de cunoștințe și aplicarea acestora în diverse domenii ce implică analiza datelor. Pașii metodei implică determinarea unei surse de date, ce conține un set de entități și aplicarea analizei semantice și analizei clusterizate pentru a extrage un set de atribute pentru fiecare entitate. Graful de cunoștințe este generat prin vectorizarea fiecărei entități din set, proces folosit pentru determinarea contextului entității, relațiile și incidențele relațiilor acesteia, ce vor forma mai departe graful. Dezavantajul acestei invenții este că nu ia în considerare descrierea text aferentă unui nod grafului, ci modelează reprezentările exclusiv prin prisma muchiilor dintre nodurile din graful de cunoștințe.

Brevetul CN107391623B (<https://patents.google.com/patent/CN107391623B/en>) prezintă o metodă de reprezentare a grafurilor de cunoștințe ce agregă informații din mai multe domenii. Pașii metodei includ alegerea unor entități etichetate anterior dintr-o bază de cunoștințe și selectarea unor corpusuri existente care sunt corelate cu entitățile selectate pentru a forma o bază de cunoștințe din mai multe surse. Pasul următor implică învățarea unei reprezentări vectoriale prin îmbinarea cunoștințelor multi-context, prin modelul DBALSTM: depth D + bidirectional B + attention A + LSTM [Wang, J., Yu, L.-C., Lai, R. K., & Zhang, X. (2016). Dimensional sentiment analysis using a regional CNN-LSTM model. Paper published in the Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2016)]. Arhitectura LSTM (Long Term Short Term Memory) Bidirecțională se folosește atât de contextul anterior, cât și de cel viitor, prin procesarea unei secvențe de text din ambele capete. Dezavantajul acestei metode este că modelul DBALSTM este inferior arhitecturilor de tip Transformer, spre exemplu BERT. De asemenea, reprezentarea construită din graful de cunoștințe este bazată pe TransE [Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In Neural Information Processing Systems (NIPS)], care nu este potrivită în cazul grafurilor cu mai multe relații de același tip având aceeași sursă.

Brevetul WO2020000848A1 (<https://patents.google.com/patent/WO2020000848A1/en>) descrie o metodă de generare automată a grafurilor de cunoștințe și un sistem care o poate folosi asupra corpusurilor masive și nestructurate. Se folosește un clasificator antrenat pentru a prezice relațiile; secvențele unitare ce conțin informații similare sunt unite între ele, și mai departe un graf de cunoștințe unificat global este creat. O limitare a prezentei invenții o reprezintă lipsa extensibilității metodei, întrucât doar 3 surse de date sunt folosite pentru informațiile legate de entități: enciclopedii, site-uri cu informație deschisă tuturor, sau o parte a bazelor de cunoștințe relevante domeniilor.

Brevetul CN111259164A (<https://patents.google.com/patent/CN111259164A/en>) descrie o metodă și un sistem de vizualizare interactivă a grafurilor de cunoștințe. Metoda presupune reprezentarea fiecărui nod dintr-un graf de cunoștințe sub forma unui simbol vizual și maparea relațiilor dintre noduri sub forma unor muchii direcționate, astfel încât grafurile de cunoștințe să aibă o reprezentare finală grafică mai ușor de urmărit și înțeles. Acest brevet nu prezintă o metodă îmbunătățită de învățare a reprezentării pentru grafurile de cunoștințe, ci le folosește drept intrare pentru generarea vizualizărilor.



Brevetul US20190370337A1 (<https://patents.google.com/patent/US20190370337A1/en>) prezintă un mecanism de generare a unei reprezentări vectoriale a conținutului unui document în limbaj natural. Metoda analizează, pentru început, structura documentului și identifică dependențele și relațiile dintre elementele din document și documentul inițial. O structură de dependențe este generată, iar muchiile ei reprezintă relația dintre una sau mai multe elemente structurale ale documentului și documentul inițial. Drept limitare, această metodă se raportează doar la relațiile dintre entitățile din document și nu realizează o reprezentare vectorială în profunzime a documentului primit.

### **Prezentarea sintetică a invenției**

Problema tehnică pe care invenția își propune să o rezolve vizează limitările metodelor actuale de învățare a reprezentării pentru grafurile de cunoștințe. Prin noțiunea de reprezentare vectorială a unui graf de cunoștințe se înțelege generarea unei reprezentări sub formă de vectori pentru elementele constituente ale grafului de cunoștințe. Metoda propusă primește ca intrare un model de limbă pre-antrenat și descrierile textuale ale entităților din graf, iar pe baza acestora realizează o antrenare pentru a genera o reprezentare vectorială a entităților și relațiilor. Ulterior, faza de analiză facilitează identificarea de entități similare, precum și de noi relații în graful de cunoștințe; totodată, se pot genera vizualizări interactive pornind de la reprezentările vectoriale obținute în urma fazei de antrenare.

Un beneficiu major al acestei invenții îl reprezintă posibilitatea de fuziona relațiile din graful de cunoștințe cu descrieri textuale în vederea obținerii unor noi reprezentări vectoriale. Metoda propusă poate fi aplicată inclusiv a posteriori pentru noi noduri adăugate în graful de cunoștințe și include posibilitatea de a reprezenta vectorial inclusiv entități care nu au o descriere textuală asociată. Metodă propusă este exemplificată pe un studiu de caz modelând entități din literatura română.

### **Avantaje**

Metoda conform invenției prezintă următoarele avantaje:

- Generarea reprezentărilor vectoriale ale unei entități se poate face folosind descrierea textuală asociată entității și a reprezentării vectoriale derivate din graful de cunoștințe;
- Generarea reprezentărilor vectoriale ale unei entități se poate realiza chiar dacă entitatea nu are asociată inițial o descriere textuală;
- Posibilitatea de alocare a unui reprezentări vectoriale nodurilor noi din graful de cunoștințe care nu existau la momentul calcului vectorilor, dacă aceste entități au o descriere textuală;
- Graful de cunoștințe permite mai multe tipuri de relații între aceleași două entități, folosind o proiecție a reprezentării vectoriale a entităților, în funcție de tipul relației;
- Descrierea textuală a unei entități influențează diferit reprezentarea finală a entității în funcție de tipul acesteia - clasa asociată în graful de cunoștințe.

### **Prezentarea figurilor**

În continuare, invenția va fi descrisă în detaliu, cu referire la Figura 1, care prezintă schema bloc a metodei propuse care utilizează graful de cunoștințe și descrierile textuale ale entităților în vederea calcului reprezentărilor vectoriale ale entităților și relațiilor, urmate de identificarea de entități similare și de noi relații, precum și vizualizarea reprezentărilor vectoriale rezultate. Figura 2 introduce o modelare formală a transformărilor care au loc la nivel de nod, respectiv la nivel de relație. Ulterior, Figura 3 introduce un studiu de caz pentru o înțelegere mai bună a contextului de exploatare a metodei propuse în domeniul literaturii române, pornind de la un graf de cunoștințe parțial care include descrieri textuale pentru entități.

### Descrierea detaliată a invenției

Metoda propusă presupune codificarea entităților și a relațiilor dintr-un graf de cunoștințe sub forma unor vectori care să permită calcularea similarității între entități diferite, dar și inferența eventualelor relații lipsă. Reprezentarea vectorială a unei entități depinde de textul asociat acesteia, de tipul ei, cât și de relațiile acesteia cu alte entități din graful de cunoștințe.

Presupunem că metoda are la dispoziție două intrări conform Figurii 1: un model de limbă pre-antrenat (1) și un graful de cunoștințe (2) cu descrierile textuale ale nodurilor.

Graful de cunoștințe [Amit Singhal. 2012. *Introducing the Knowledge Graph: things, not strings*. Google Blog. <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>] conține entități de diferite tipuri (noduri) și relații (muchii în graf). Tipurile de entități sunt modelate folosind clase definite folosind OWL [McGuinness, L. and van Harmelen, F. (2004). *OWL Web Ontology Language Overview -W3C Recommendation*. Paper published at <http://www.w3.org/TR/2004/REC-owl-features-20040210/> (2004)] sau RDFS [Allemang, D. and Hendler, J. (2008). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Book published at Morgan Kaufmann (2008)]. Suplimentar, în cadrul ontologiei pot exista instanțe ale unei clase cu proprietăți aferente definite. Entitățile au fiecare o posibilă descriere textuală, relațiile dintre entități sunt de diverse tipuri, iar metoda propusă de codificare vectorială va trebui să țină cont de acestea.

Descrierea textuală a entității poate fi transformată într-o reprezentare vectorială folosind modele de limbă pre-antrenate bazate pe arhitectura de tip Transformer [Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. Paper presented at the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA], cum ar fi BERT [Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171-4186)], sau alte modele semantice de reprezentare vectorială a textelor (de ex., Doc2Vec [Lau, J. H. and Baldwin, T. (2016). *An empirical evaluation of doc2vec with practical insights into document embedding generation*. In *Proceedings of ACL Workshop on Representation Learning for NLP*. (2016)]). Modelele de limbă bazate pe Transformer sunt de preferat însă, acestea furnizând reprezentări

contextualizate superioare și sunt folosite în ultimul timp în majoritatea problemelor legate de prelucrarea limbajului natural.

Metoda presupune următoarele faze reliefate în Figura 1:

1. *Faza de antrenare*, centrată pe calculul reprezentărilor vectoriale ale entităților și a relațiilor (3);
2. *Faza de analiză bazată pe noul spațiu de reprezentare*, în cadrul căreia sunt identificate entități similare și noi relații (4), respectiv sunt afișate reprezentările vectoriale ale entităților (5).

Reprezentarea vectorială a textului asociat entității  $e$  va fi notată cu  $text(e)$ , vectorul rezultat având dimensiunea fixată  $b$ , care depinde de modelul ales de reprezentare a textului. Pentru proiectarea vectorului  $text(e)$  în spațiul de reprezentare a entităților, va fi folosită o înmulțire cu o matrice  $M_{type(e)}$  de dimensiune  $d \times b$ . Această matrice depinde de tipul entității  $e$ , notat cu  $type(e)$ . Fiecare entitate va avea asociat și un vector de dimensiune  $d$  inițializat aleatoriu și modificat în timpul antrenării,  $emb(e)$ . Astfel, reprezentarea finală a unei entități  $e$  va fi calculată prin formula următoare, unde  $\sigma$  poate fi orice funcție de activare (sigmoid, tanh, Relu, sau chiar funcția identitate):

$$v(e) = \sigma(\sigma(M_{type(e)} * text(e)) + emb(e))$$

Acest mod de generare a unei reprezentări are avantajul de a putea calcula vectori și pentru entități noi, necunoscute în etapa de antrenare, folosind descrierea asociată acesteia, dar și pentru entități care au o reprezentare  $emb(e)$ , dar nu au și o descriere.

Fiecare tip de relație dintre entități  $r$  va avea asociată o matrice de transformare  $A_r$  de dimensiune  $d \times d$  și un vector  $emb(r)$  de dimensiune  $d$ . Pentru orice triplet  $\langle h, r, t \rangle$  din graful de cunoștințe, unde  $h$  și  $t$  sunt cele două entități legate prin relația  $r$ , modelul propus va fi antrenat să prezică entitatea țintă  $t$  folosindu-se de reprezentările vectoriale ale lui  $h$  și  $t$ , precum și de transformarea definită prin relația  $r$ .

Toate transformările anterioare sunt reprezentate vizual în cadrul Figurii 2.

Reprezentările și toate matricele de transformare folosite sunt învățate prin backpropagation folosind următoarea funcție obiectiv, pentru fiecare triplet  $\langle h, r, t \rangle$  din graful de cunoștințe:

$$|A_r * v(h) + emb(r) - A_r * v(t)|_L$$

În faza de analiză se folosesc vectorii și matricele de transformare calculate în etapa de antrenare. Pentru recomandarea de entități similare se poate folosi o metrică de similaritate pe vectorii calculați deja, cum ar fi similaritatea cosinus. Similaritatea poate fi calculată atât pe vectorul final, cât și pe reprezentarea descrierilor textuale asociate celor două entități.

Pentru identificarea de noi relații într-un graf de cunoștințe se pot selecta capetele eventualei relații dintre entitățile din graf și se aplică funcția obiectiv din cadrul etapei de antrenare pentru fiecare tip posibil de relație care ar putea exista între cele două tipuri de entități. Valoarea obținută poate fi comparată cu distribuția valorilor pentru relațiile existente deja în graf, pentru a stabili un prag de la care relația nouă poate fi adăugată în graful de cunoștințe.

Totodată, metoda propusă include vizualizarea și explorarea grafului de cunoștințe modelat în noul spațiu de reprezentare vectorială. Pentru început, o viziune de ansamblu a întregului graf este oferită, în care entitățile pot fi grupate în diverse categorii, spre exemplu după tip/clasă aferentă, pentru simplificarea reprezentării grafice. Mai mult, având în vedere dimensionalitatea reprezentării vectoriale latente, se poate aplica o descompunere de tip PCA [Wold, S., Esbensen, K., Geladi, P. (1987). *Principal component analysis. Paper published in Chemometrics and Intelligent Laboratory Systems, 2 (1–3) (1987), pp. 37-52.*] pentru reducerea dimensionalității sau t-SNE [van der Maaten, L. and Hinton, G. E. (2008). *Visualizing data using t-SNE. Paper published in J. Mach. Learn. Research 9, 2579–2605 (2008)*], proiectând întregul graf în două dimensiuni. Ulterior, pe graf se pot realiza operații de zoom-in, respectiv zoom-out, care extind, respectiv comprimă, nivelul de granularitate la care sunt afișate entitățile. Totodată, pot fi identificate și vizualizate cele mai apropiate  $k$  entități din noul spațiu de reprezentare. Pentru restrângerea suplimentară a entităților vizualizate, se pot aplica filtre în funcție de tipul de legături dintre entități sau diferite proprietăți ale entităților. Interacțiunea cu graful se realizează prin identificarea de relații între entități, pentru aceasta fiind necesară selectarea a două entități, respectiv prin solicitarea găsirii entităților similare pentru o entitate selectată.

Figura 3 ilustrează un studiu de caz în care metoda propusă este aplicată în domeniului literaturii române, folosind drept intrare modelul de limbă pre-antrenat RoBERT [Masala, M., Ruseti, S., & Dascalu, M. (2020). *RoBERT – A Romanian BERT Model. In 28th Int. Conf. on Computational Linguistics (COLING) (pp. 6626–6637). Barcelona, Spain (Online): ACL*] disponibil pentru limba română. Graful de cunoștințe are 4 clase marcate cu bold și anume:

- Scriitor, clasă care încapsulează descrierile scriitorilor din Dicționarul General al Literaturii Române (DGLR) publicat de Academia Română;
- Eveniment Literar, clasă care încapsulează descrierea de evenimente literare, ce implică scriitori și reviste literare din perioada 1944-2000, din corpusul Cronologiei Vieții Literare Românești (CVLR) publicat de Academia Română;
- Publicație, clasă care încapsulează multiple subtipuri precum volume de poezii, nuvele sau romane ale scriitorilor români;
- Revistă Literară, sub-clasă a clasei Publicație.

În cadrul grafului, relațiile dintre entități și proprietățile acestora sunt descrise conform Resource Description Framework (RDF) [Beckett, D., & McBride, B. (2004). *RDF/XML syntax specification (revised). W3C recommendation, 10(2.3)*]. Suplimentar față de clasele anterioare, Figura 3 introduce și instanțe precum: Nichita Stănescu pentru clasa Scriitor, volumul “Sensul Iubirii” pentru clasa Publicație, Gazeta Literară instanță a clasei Revistă Literară, sau data “31 mai 1962” instanță pentru clasa Eveniment Literar. Instanțele specifice ale diverselor clase pot avea relații între ele pe baza cărora acestea interacționează. Spre exemplu, o instanță a unui Scriitor publică o instanță a unei Publicații. Instanța unui Scriitor poate participa în cadrul unei instanțe Eveniment Literar. Clasa principală a grafului de cunoștințe din Figura 3 o reprezintă Scriitorul, aflat în relație cu clasa Revistă Literară, relația fiind definită drept: “Scriitorul  $x$  este menționat în Revista Literară  $y$ ”.



Pornind de la graful de cunoștințe din Figura 3, coroborat cu descrierile textuale din DGLR și CVLR aferente entităților furnizate, o multitudine de aplicații practice potențiale pot fi desprinse aplicând metoda propusă în cadrul brevetului. Spre exemplu, șabloane pot fi identificate reliefând similarități între scriitorii din literatura română (spre exemplu, Ion Barbu este un poet român, considerat scriitor canonic, care a activat la începutul secolului 20, similar cu Tudor Arghezi, fiind contemporani, și el scriitor canonic), asocieri între reviste literare și între evenimente apărute de-a lungul timpului, cuplate cu vizualizări interactive pornind de la reprezentările vectoriale în noul spațiu de reprezentare.





## Revendicări

1. Metodă de calcul a reprezentărilor vectoriale ale entităților și relațiilor dintre ele, cuprinzând o fază de antrenare și o fază de analiză a informațiilor din graful de cunoștințe folosind noul spațiu de reprezentare, **caracterizată prin aceea că**, în *faza de antrenare*, se aplică un algoritm bazat pe rețele neurale pentru calcul de reprezentări într-un graf de cunoștințe care cuprinde entități sau noduri cu descrieri textuale în limbaj natural și relații dintre noduri (muchii în graf),

2. Metodă de calcul **conform revendicării 1, caracterizată prin aceea că** algoritmul aplicat cuprinde o primă etapă de calcul al reprezentării descrierilor folosind modele de limbă sau alte modele semantice aplicate asupra descrierilor textuale, iar într-o a doua etapă se calculează reprezentarea unui nod prin combinarea descrierii textuale vectorizate cu reprezentarea nodului modificat la antrenare; într-o a treia etapă se determină proiecția entităților dintr-un triplet de forma <nod sursă, relație, nod destinație> în spațiul de reprezentare al relației dintre noduri; într-o a patra etapă se calculează distanța dintre nodul sursă și relație raportat la nodul destinație; într-o a cincea etapă se actualizează ponderile pe baza distanțelor calculate prin backpropagation,

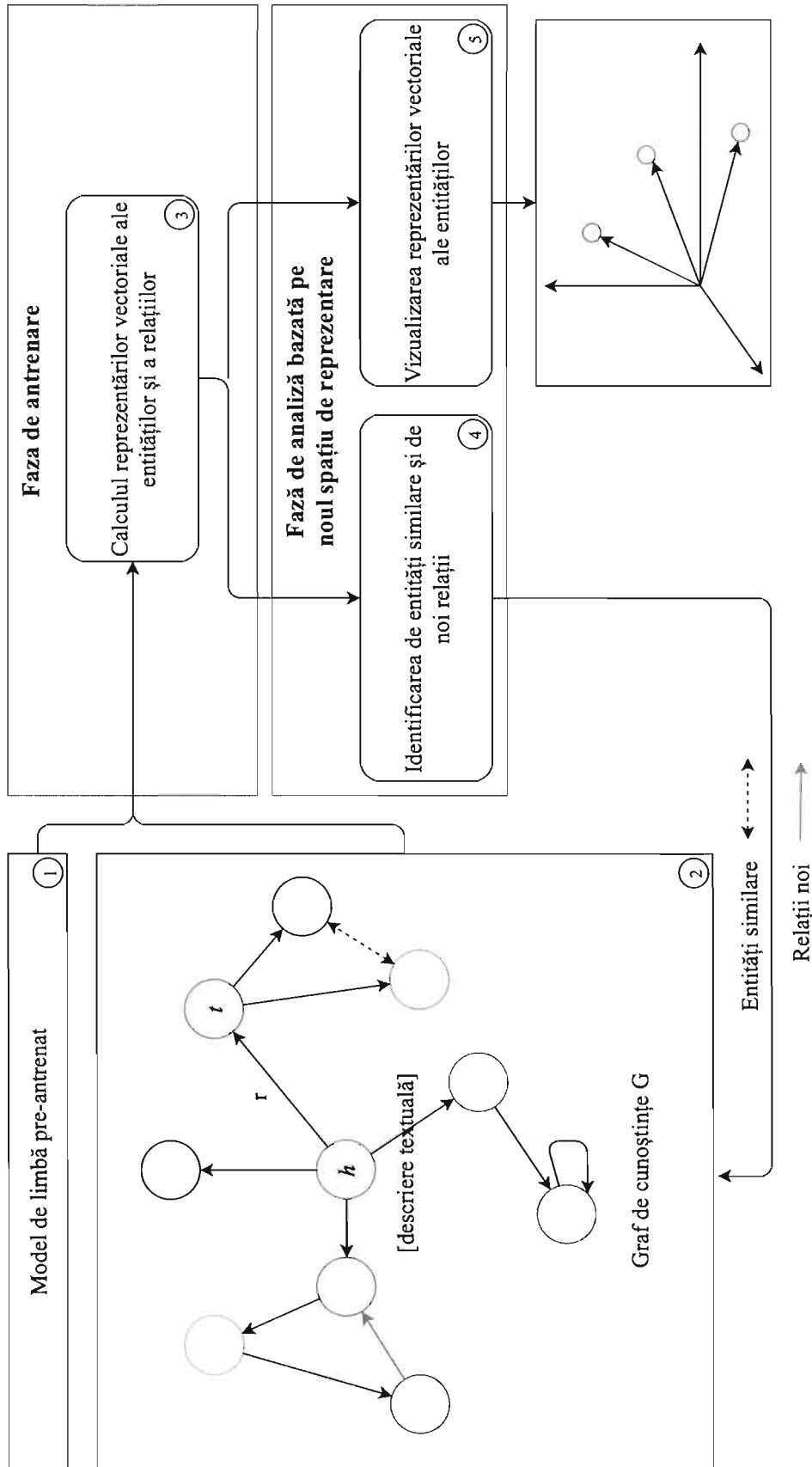
3. Metodă de calcul **conform revendicării 1, caracterizată prin aceea că** repetarea etapelor 2-5 ale algoritmului se realizează până la finalizarea antrenării și convergența rețelei, și se trece la o *fază de analiză bazată pe noul spațiu de reprezentare* în care sunt recomandate entități similare folosind o metrică de similaritate aplicată pe reprezentările vectoriale finale ale entităților sau pe reprezentările descrierilor entităților, sunt identificate posibile noi relații în scopul minimării distanței dintre primul nod și relație, raportat la al doilea nod în spațiul de reprezentare al relației, și sunt vizualizate reprezentările vectoriale ale entităților din graful de cunoștințe.

4. Metodă **conform revendicării 1, caracterizată prin aceea că** descrierea textuală a entităților, reprezentată prin modele de limbă pre-antrenate bazate pe arhitectura de tip Transformer, cum ar fi BERT, sau alte modele semantice de reprezentare vectorială a textelor (spre ex., Doc2Vec), este combinată, printr-o proiecție într-un alt spațiu vectorial influențată de tipul entității, cu o reprezentare vectorială a entității derivată din graful de cunoștințe.

5. Metodă **conform revendicării 1, caracterizată prin aceea că** reprezentările vectoriale ale entităților legate printr-o relație sunt proiectate în spațiul aferent relației înainte de calculul distanței dintre nodul sursă, relație și nodul destinație.

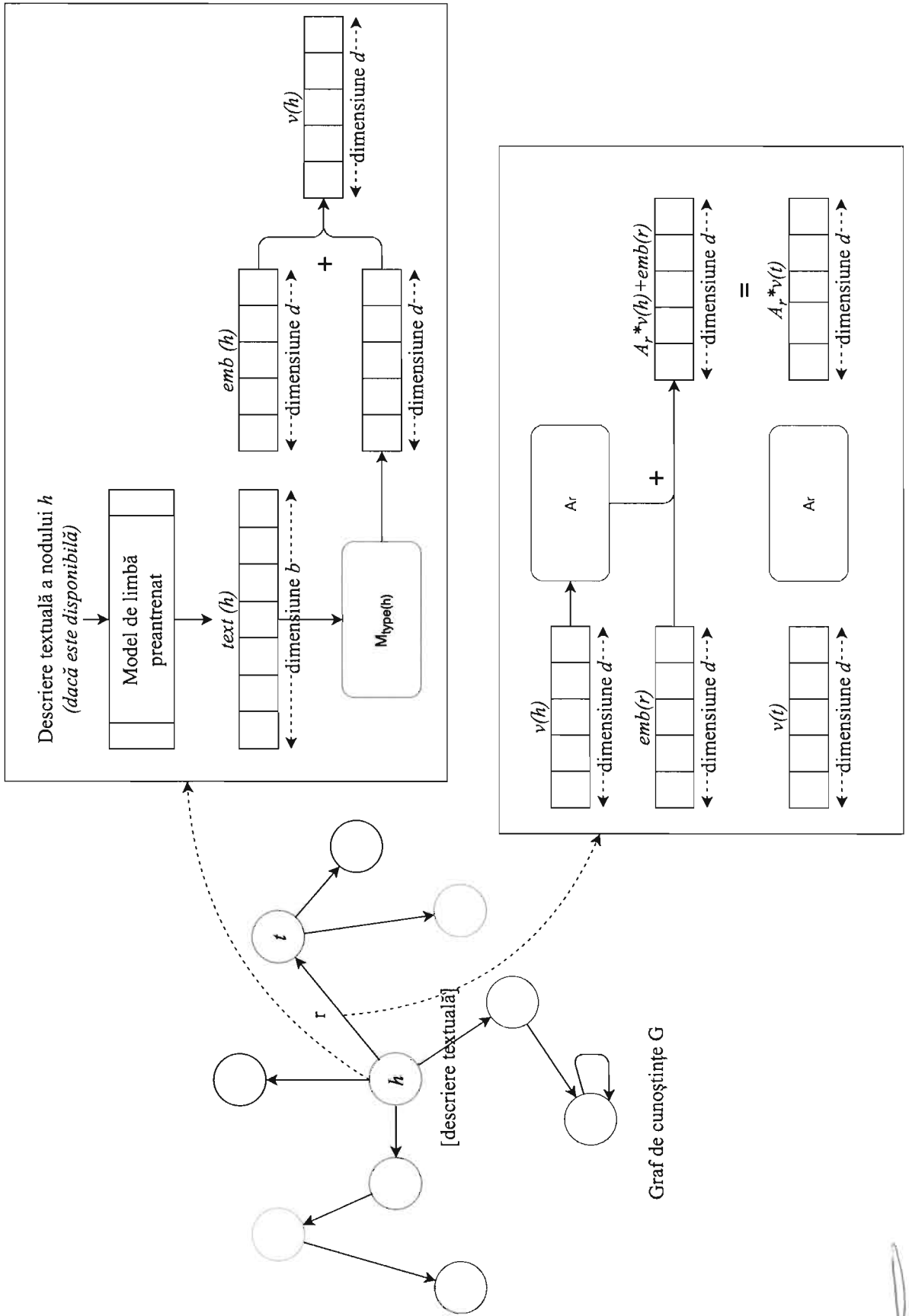


Figura 1. Schema bloc a metodei de învățare automată de reprezentări vectoriale în frazuri de cunoștințe pornind de la modele de limbă



*[Handwritten signature]*

Figura 2. Modelarea formală a transformărilor metodei la nivel de nod și de relație



Graf de cunoștințe G

*[Handwritten signature]*

Figura 3. Exemplu de graf de cunoștințe și descrieri textuale în domeniul literaturii române.

