Universitatea Transilvania din Brașov

Universitatea Transilvania din Brașov
FACULTATEA DE INGINERIE ELECTRICĂ ȘI ȘTIINȚA CALCULATOARELOR

ȘCOALA DOCTORALĂ INTERDISCIPLINARĂ

FACULTATEA DE INGINERIE ELECTRICĂ ŞI ŞTIINȚA CALCULATOARELOR

Ing. Dipl. ADRIAN M.P. BRAŞOVEANU

Sisteme Inteligente în Rețele Semantice

Intelligent Systems in Semantic Networks

REZUMAT / ABSTRACT

Conducător ştiințific / Supervisor

Prof.dr.mat. RĂZVAN ANDONIE

BRAȘOV, 2021

D-lui(D-nei).....

COMPONENȚA

Comisiei de doctorat

Numită prin ordinul Rectorului Universității Transilvania din Braşov
Nr. ..... din .....

PREŞEDINTE:	- CONF. UNIV. DR. CARMEN GERICAN Universitatea Transilvania din Braşov
CONDUCĂTOR ŞTIINIFIC:	- PROF. UNIV. DR. MAT. RĂZVAN ANDONIE Universitatea Transilvania din Braşov
REFERENȚI:	- PROF. UNIV. DR. DIANA INKPEN Universitatea din Ottawa, Canada
	- PROF. UNIV. DR. GHEORGHE ȘTEFAN Universitatea Politehnică din Bucureşti
	- PROF. UNIV. DR. LIVIU P. DINU Universitatea din Bucureşti

Data, ora şi locul susținerii publice a tezei de doctorat: ....., ora ....., sala .....

Eventualele aprecieri sau observații asupra conținutului lucrării vor fi transmise electronic, în timp util, pe adresa brasoveanu.adrian@unitbv.ro.

Totodată, vă invităm să luați parte la şedința publică de susținere a tezei de doctorat.

Vă mulțumim.

CUPRINS (lb. română)

CONTENT

CAPITOLUL 1

INTRODUCERE

1.1 FUNDAL

Sistemele inteligente au fost mult timp un proxy pentru inteligența artificială (AI). AI este de obicei definită ca fiind căutarea inteligenței asemănătoare omului sau super-umană în mașini. Termenul Învățare Automată (ÎA sau ML) cuprinde unul dintre diverse școli de AI care au devenit proeminente în anii 1980 și ulterior au ajuns să domine campul. Ideea sa principală este că mașinile pot învăța din datele pe care le alimentează.

Există și alte școli de AI, printre care și Semantic Web (SW), școală cunoscută și sub numele de Grafuri de Cunoaștere (GC). Sunt implementate grafurile de cunoaștere (GC) folosind un format triplet (de exemplu, subiect-predicat-obiect) care permite utilizarea de limbaje de interogare expresive pentru găsirea faptelor, precum și raționamente automate pentru a deduce date noi. Scopul final al procesării limbajului natural (PLN sau NLP) este decodarea limbii umane în diversele lor formate. În această teză, vedem NLP ca fiind legătură între cele două ramuri ale AI discutate în această teză: ML și SW.

1.2 CONTRIBUȚII PRINCIPALE

IA Semantică (IAS) este o abordare recentă de IA care se concentrează pe combinarea semanticii cu metode clasice de IA, cum ar fi clasificarea, gruparea sau recomandarea. Adăugând semantică, putem crește calitatea datelor simultan înlăturând abordările de tip cu cutie neagră. Propunerea de bază a IAS este aceea că indiferent de proveniența sa originală (de exemplu, text, tabel, imagine), datele pot fi procesate și stocate în formate rafinate precum cele furnizate de GC sau motoare de căutare. Aceste clustere de date deschise pot fi utilizate ulterior pentru a rezolva probleme complexe cu abordări hibride. Prin combinarea entităților extrase dintr-un GC cu sentiment și clasificatori ML, este posibil să se verifice afirmațiile dintr-o propoziție, de exemplu. Această teza examinează mai multe metode hibride permise de IAS pentru a înțelege cum să construim linii de bază pentru cercetare și producție. Apoi întreabă ce putem face pentru a îmbunătăți aceste metode hibride, deoarece se pare că fiecare componentă a sistemului poate adăuga erorile sale în stivă și poate deruta cercetătorii și dezvoltatorii.

Rezultatele prezentate aici pot fi clasificate în următoarele direcții de cercetare.

1.2.1 IAS pentru extragerea de cunoștiințe

Contribuțiile din această arie sunt legate de dezvoltarea de sisteme IAS, dar și de rolul pe care schimbări mici ale unui corpus, cum ar fi adăugarea unor entități sau relații, pot juca un rol în obținerea de rezultate mai bune.

- Prima contribuție este legată de rolul variației numelor pentru legarea entităților prin: (i) îmbunătățirea acoperirii sistemelor de dezambiguizare a entităților și (ii) folosirea de lentile care să evalueze diferite forme de suprafață ale acelorași entități.
- A doua contribuție este legată de construire unor modele independente de domeniu pentru analiza sentimentului.
- A treia contribuție este dedicată verificării faptelor și ne arată cum tehnicile precedente pot fi integrate pentru detecția știrilor false.

1.2.2 Interpretarea și explicarea IAS

Depanarea sistemelor IAS este extrem de dificilă pentru programatori deoarece nu este întotdeauna clar care componente generează erorile. Contribuțiile din această arie adresează acest neajuns.

- Prima contribuție este o metodologie generală de dezvoltare a unor metode de evaluare explicabile pentru sisteme IAS. Una dintre primele idei a fost dezvoltarea unei taxonomii a erorilor din dizambiguizarea de entități.
- A doua contribuție este extensia naturală a contribuției precedente, deoarece filozofia și codul din spatele taxonomiei au fost extinse într-un framework numit Orbis, care este utilizat pentru a evalua și explica rezultate din mai multe domenii de cercetare, cum ar fi dizambiguizarea entităților, extragere de relații și extragere de conținut din forumuri.
- Ultima contribuție este mai mult teoretică, fiind un sondaj comprimat legat de rolul vizualizării în explicarea și interpretarea IAS.

1.3 ORIGINI

Capitolul 1 acoperă introducerea și, prin urmare, nu este dedicat niciunei publicații.

Capitolul 2 descrie arhitectura de bază a sistemelor IAS. Studiul de caz prezentat este un sistem creat pentru afișarea indicatorilor de turism care a fost publicat în două articole de revistă în Semantic Web (IF = 3.524) [BSS + 17] și Journal of IT and Tourism (IF = 2,95) [SOBS16] și a unui articol de conferință la ENTER 2015 [SBÖ15]. Ultima secțiune a capitolului oferă o vizualizare la nivel înalt a arhitecturii SAI și reprezintă o introducere pentru capitolele următoare.

Capitolul 3 prezintă trei contribuții legate de construirea sistemelor IAS.

Capitolul 3.1 discută despre contribuțiile la dizambiguizarea entităților, și anume lentilele și variantele de nume. O entitate poate avea mai multe nume, o problemă pe care o numim variația

numelui și, prin urmare, este important ca un sistem de dizambiguizare să poată extrage toate aceste nume. O posibilitate este extragerea de nume suplimentare din mai multe GC. O altă posibilitate este de a crea algoritmi care să calculeze aceste variante de nume algoritmic. Aceste idei au fost publicate inițial într-o publicație de conferință la WIMS 2018 [WKB18] și LDK 2019 [WBKN19]. Este prezentată și o metodă prin care scorurile dizambiguizării pot fi calculate atunci când luăm în considerare și varianța numelui. Metoda implică crearea unor lentile care se concentrează pe o singură proprietate (de exemplu, mențiune, tip, link) și arată cum acest mecanism îmbunătățește rezultatele pentru unele sisteme NEL. Principalele constatări discutate în această secțiune se bazează pe două publicații de conferințe de la ACL CoNLL 2020 [BWN20] și ACM WIMS 2018 [BNW18].

Capitolul 3.2 prezintă o contribuție legată de construirea clasificatorilor afectivi specifici domeniului. Ideea principală este folosirea grafurilor de cunoștințe, embeddings, și modele de limbaj pre-instruite pentru a îmbunătăți clasificarea emoțiilor. Constatările au fost publicate într-un articol al revistei Cognitive Computation (IF = 4.307) [WSB + 21].

Capitolul 3.3 este dedicat verificării faptelor. O versiune a verificării faptelor numită detecția automată a știrilor false este examinată. O conductă NLP de bază care include entități, sentiment și relații este utilizată pentru a evalua gradele de adevăr asociate unor seturi de date bazate pe Politifact. Aceste descoperiri au fost publicate inițial într-un articol de conferință în IWANN 2019 [BA19] și apoi într-un articol de jurnal de la Neural Processing Letters (IF = 2.891) [BA20a].

Capitolul 4 este dedicat interpretabilității și explicabilității. Acestea sunt poate cele mai importante subiecte de astăzi în lumina succesului neașteptat al ML din ultimul deceniu. Dacă vrem să dezvoltăm rețele care vor diagnostica pacienții sau judeca niște cazuri legale (sau chiar oferi ajutor pentru aceste sarcini), atunci este necesar să le explicăm clar raționamentul.

Capitolul 4.1 este axat pe evaluarea explicabilă. Inițial o taxonomie a fost dezvoltată pentru a ajuta la clarificarea componentelor de dizambiguizare a entităților care declanșează anumite tipuri de erori. Mai târziu, această taxonomie a fost transformată într-un framework care ajută și la vizualizarea diferitelor erori. Taxonomia a fost prezentată la LREC 2018 [BRK + 18]. Orbis a fost introdus la SEMANTICS 2018 [OKBW18]. Un articol asociat despre îmbunătățirea standardelor de aur a fost publicat la RANLP 2019 [WBKN19].

Capitolul 4.2 discută rolul vizualizării în explicarea sistemelor de IA. Un sondaj este realizat pentru a identifica tendințele în domeniu. Secțiunea se bazează pe un articol publicat în conferința IEEE IV2020 [BA20b].

Capitolul 5 formulează concluziile și, prin urmare, citează unele dintre aceste articole.

1.4 STRUCTURA

Teza este structurată în jurul celor două mari direcții de cercetare.

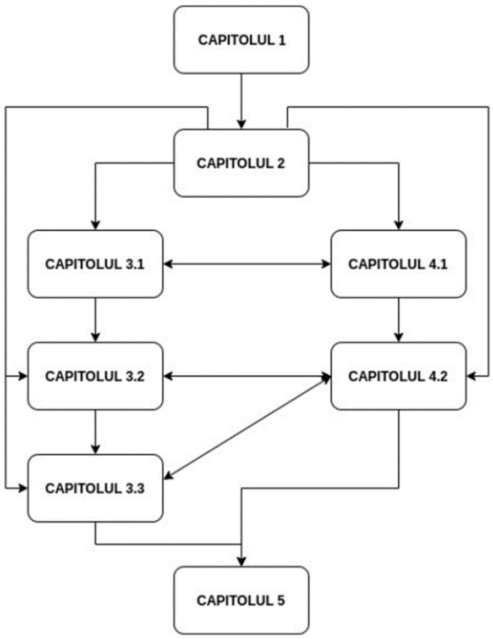

Figura 1.1 Structura tezei.

Figura prezintă legăturile dintre diferitele capitole și subcapitole, numite secțiuni. De exemplu, Capitolul 2 prezintă bazele GC și are legături cu majoritatea secțiunilor următoare. Restul capitolelor construiesc pe baza materialului precedent. În unele cazuri, legături între secțiuni din capitole diferite pot fi și ele observate, cum ar fi capitolele 3.1 și 4.1 care discută dizambiguizarea entităților din două puncte de vedere diferite.

CAPITOLUL 2

BAZELE IA SEMANTICE

2.1 O SCURTĂ ISTORIE A GRAFURILOR DE CUNOȘTINȚE

Există multe definiții ale unui GC. Cele mai multe dintre ele sunt legate de aspectul de Reprezentare a Cunoașterii (RC) al grafurilor (de exemplu, graful de proprietăți, graful direcționat, etc). Unii sunt interesați și de formalismul logic din spatele acestuia. În scopurile noastre, preferăm o definiție simplă care se bazează pe definiția lui Aidan Hogan [HBC + 20]:

Definiție. Un graf de cunoștințe (GC) este un graf care conține o reprezentare limitată a lumii reale și ale cărui noduri și margini reprezintă entități din viața reală și relațiile dintre ele.

GC pot accepta limbaje de interogare, construcții de RC (de exemplu, ontologii sau reguli) și chiar mai multe serializări (de exemplu, RDF, JSON).

Limbajele de interogare GC trebuie să fie mai expresive decât SQL, pentru că trebuie să permită atât operatori relaționali (de exemplu, uniuni), precum și operatori recursivi care pot include expresii ale căilor (de exemplu, expresii care se pot potrivi cu căile de navigare dintre noduri). SPARQL este cel mai utilizat limbaj de interogare GC.

Ontologiile, un formalism standard în RC, reprezintă entități și relații dintr-un domeniu. Acestea conțin instrucțiuni care definesc domeniul (TBox), precum și câteva exemple sau instanțe ale claselor definite (ABox). Ontologiile sunt definite folosind expresii logice și permit adesea raționamente asupra datele colectate. Cel mai utilizat format pentru construirea ontologiilor de astăzi este OWL. Motoarele de raționare moderne folosesc adesea un subset al OWL, cum ar fi OWL DL, pentru a-și exprima construcțiile logice.

Au fost necesare mai multe formate de serializare deoarece RDF, modelul original de metadate pentru SW scris în XML, a fost dificil de utilizat. Formalizarea timpurie s-a concentrat pe tripleți de genul subiect-predicat-obiect (de exemplu, formatele N3, N-Triples), în timp ce mai recent serializările acceptate includ și JSON (de exemplu, JSON-LD). Unele GC pot fi, de asemenea create fără astfel de construcții (de exemplu, bazate pe baze de date), dar vor fi folosi și un formalism limitat.

Accesul la GC care susțin aceste construcții este de obicei primul pas spre dezvoltarea de aplicații semantice.

O aplicație clasică este o platformă pentru accesarea și vizualizarea GC. Astfel de platforme sunt în general denumite Platforme de date conectate (LDP) și tind să respecte principiile de publicare a datelor

legate descrise de Sir Timothy Berners-Lee și mai mulți dintre colaboratorii săi, principii numite Linked Data. Ideea principală din spatele acestui set de practici a fost publicarea de seturi de date online folosind identificatori unici pentru entități. Acest lucru le-a permis interogarea acestor seturi de date prin limbaje precum SPARQL, dar a permis și legarea lor de alte seturi de date, extinzând astfel graficul de date deschise conectate numit Linked Open Data (LOD). GC precum DBpedia și Wikidata sunt hub-uri centrale în graficul LOD.

Aplicații mai complexe pot fi, de asemenea, imaginate și includ elemente de procesare a limbajului natural și învățare automată. Astfel de sisteme folosesc în general o suită de instrumente de extragere a cunoștințelor (IE) sau vor fi ele însele Instrumente IE. Una dintre cerințele principale pentru acestea va fi extragerea datelor (de exemplu, text, entități, sentiment) de pe web sau documente (de exemplu, fișiere PDF) și afișarea într-un format ușor de interpretat de oameni.

2.2 CONSTRUIREA DE APLICAȚII SEMANTICE FĂRĂ IA

2.2.1 Accesarea datelor pe baze ontologice

Grafurile de Cunoștințe pot de asemenea reprezenta date din bazele de date clasice. În unele cazuri, bazele de date pot fi accesate automat într-un format virtual GC prin instrumente de acces la date bazate pe ontologie (OBDA) [CCK + 17]. GC virtual poate fi creat prin mapări care vor specifica corespondența dintre entitățile bazei de date și conceptele ontologice. Unele instrumente OBDA pot, de asemenea materializa GC mapate, ceea ce înseamnă că pot crea arhive cu toate triplurile din GC. Ontop [CCK + 17] este unul dintre cele mai bune instrumente OBDA.

Baza de date TourMIS [Wöb03] a fost transformată într-un GC folosind metoda OBDA și Ontop. Procesul de construcție este prezentat în [SOBS16].

2.2.2 Linked Data și tablouri de bord

Una dintre cele mai importante aplicații ale proiectului ETIHQ a fost construcția unui tablou de bord pentru prezentarea indicatorilor statistici. Acest tablou de bord a fost construit folosind platforma webLyzard care integrează servicii de date și vizualizări. Construcția tabloului de bord este descrisă în [BSS + 17].

O captură de ecran a tabloului de bord rezultat este prezentată în Figura 2.1. Cum bine se poate vedea din captura de ecran, graficele ne permit să tragem rapid concluzii legate de datele afișate. Se vede cu ușurință că există un fel de sezonalitate a datelor, cu vârfuri clare în timpul verii și minime locale pe durata iernii. Acest lucru se datorează faptului că niciuna dintre capitalele vizualizate nu era o destinație de iarnă.

După cum se vede, chiar și dacă tablourile de bord asociate includ informații semantice, nu putem răspunde la întrebări complicate uitându-ne la diagrame statistice. De exemplu, ce a provocat vârfurile? De ce aceste vârfuri diferă între țări sau ani? Datele în sine nu vor oferi destule indicii, cu excepția cazului în care aceste detalii suplimentare sunt prezentate prin multe căutări pe Internet.

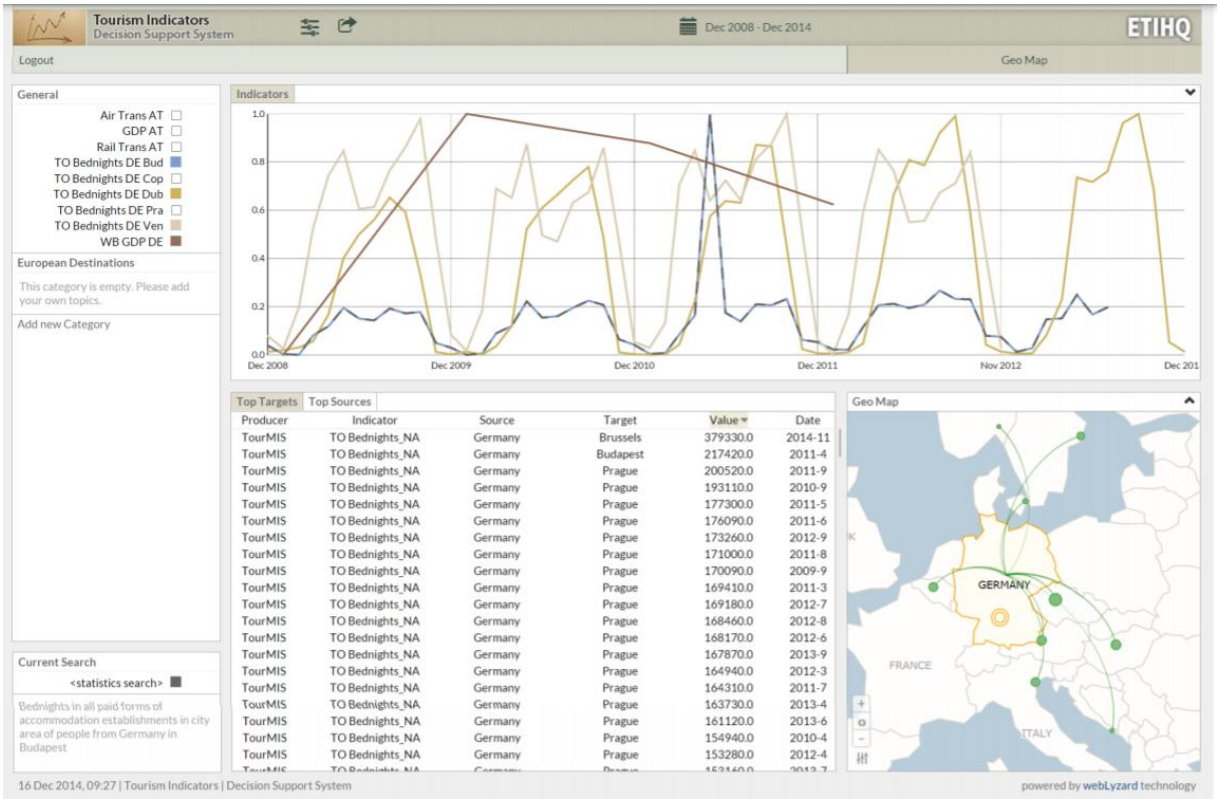

Figura 2.1 Interfața ETIHQ. Reprodusă după [BSS+17].

Pentru a obține răspunsuri mai bune, este important să introducem noi tipuri de date. Am putea, de exemplu, să analizăm articole de știri publicate în timpul perioadei examinate. Acest lucru va necesita arhitecturi mai complexe, precum cele descrise în paginile următoare.

2.3 ARHITECTURA IAS

Figura 2.2 include elementele componente ale unei arhitecturi de IAS. Acestea includ: (i) documente exprimate în limbaj natural, (ii) conducte NLP, (iii) GC și (iv) aplicații.

2.3.1 Documente în limbaj natural

Indiferent de proveniența lor (de exemplu, articole de presă, rețele sociale, forumuri sau audio / video), majoritatea informațiilor disponibile online pot fi transformate în text natural (NL). Pentru a accesa textul, trebuie să-l curățăm și să-l culegem.

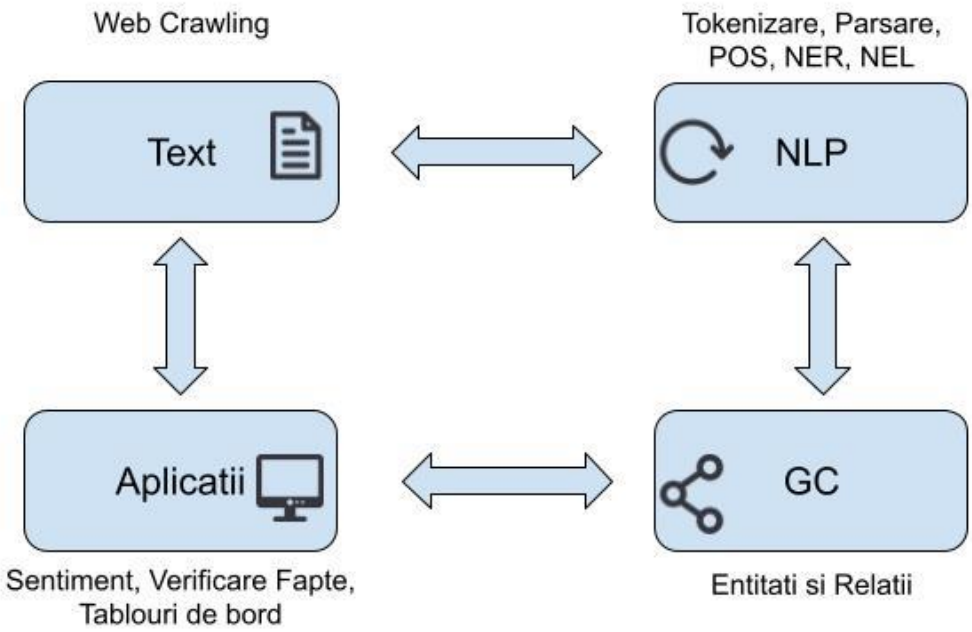

Figura 2.2 Arhitectura NLP

2.3.2 Procesarea limbajului natural

Sarcina principală a NLP a fost întotdeauna să extragă informații semnificative din documente în limbaj natural (NL). Uneori acest lucru însemna generarea de traduceri și, uneori, adăugarea de adnotări pentru a identifica diferite entități sau evenimente. Există instrumente dedicate pentru sarcinile NLP, dar multe dintre aceste instrumente pot fi accesate prin intermediul interfețelor linie de comandă (CLI) sau API-urilor REST și înlănțuite împreună pentru a forma conducte NLP. Acest lucru este necesar, deoarece adesea aceste sarcini depind una de alta.

2.3.3 Grafuri de cunoștințe

În funcție de sistemul pe care dorim să îl construim, poate fi nevoie de unul sau mai multe grafuri de cunoștințe. Dacă sistemul are nevoie pur și simplu de legăturile dintre entități, atunci poate fi suficient să furnizăm pur și simplu aceste link-uri. Dacă sarcinile care trebuie rezolvate sunt mai complicate și implică, de asemenea, găsirea entităților numite sau completarea informațiilor lipsă în GC (de exemplu, umplerea sloturilor), atunci este posibil să trebuiască să furnizăm infrastructură pentru efectuarea acestor operațiuni suplimentare.

2.3.4 Aplicații

Unele aplicații complexe sunt construite în jurul analizei sentimentului sau verificării faptelor. Cele mai multe dintre vizualizări sau tablouri de bord, precum și LDP-urile vor sta pe acest strat.

Deși acest capitol a oferit o scurtă istorie a celor mai importante descoperiri legate de GC, nu a făcut același lucru pentru ML. Aceasta este pur și simplu pentru că unele aspecte ale ML sunt discutate în detaliu în capitolele următoare în scurtele secțiuni de fundal pentru fiecare dintre problemele discutate. Câteva alte texte pot fi consultate pentru a găsi noi informații. Un sondaj legat de aplicațiile Deep Learning pentru descoperirea științifică poate fi găsit în [RS20]. Mai aproape de subiecte abordate în această teză, sondaje despre aplicațiile ML și DL pentru NLP pot fi găsite în [YHPC18] și [OMK21].

2.4 MODELE DE LIMBAJ ȘI IAS

GC nu sunt singura opțiune pentru decodificarea semnificațiilor. Un argument serios poate fi realizat și pentru modelele de limbaj modern (LM) precum Transformers. Ideea principală implementată de arhitectura Transformer a fost că este suficient să ne concentrăm pur și simplu asupra atenției pentru a ajunge la rezultate bune [VSP + 17]. Un set de capete de atenție multiple pare a fi tot ceea ce este necesar pentru a oferi performanță medie pentru sarcinile NLP. Acest mecanism oferă modelului posibilitatea de a interpreta o varietate de subspații de reprezentare la diverse poziții. Aceasta înseamnă că mai multe matrice de greutate pot fi procesate în paralel, codificând diferite cuvinte sau fraze dintr-un document. Ieșirile sunt introduse în perechi de codificatoare și decodificatoare, în funcție de model. Codificatoarele și decodificatoarele pot fi utilizate pentru diferite sarcini (de exemplu, codificatoare pentru extragerea entităților și decodificatoarele pentru clasificarea emoțiilor [RS20]).

Deoarece astfel de LM pot oferi performanțe bune pentru sarcini precum analiza dependențelor (DP), extragerea entităților sau răspunsul la întrebări, este normal să întrebăm dacă astfel de modele de limbaj ar trebui considerate AI semantice? Răspunsul scurt este da, unele dintre ele pot fi considerate IAS. Răspunsul lung este puțin mai nuanțat. Dacă un model de limbă își poate crea reprezentarea asupra lumii, atunci va fi un IAS. Dacă nu poate face acest lucru, atunci nu va fi IAS. Unde ar trebui să tragem linia? Sisteme inteligente care pot salva reprezentările învățate într-un format accesibil ar trebui considerate IAS, din punctul nostru de vedere. Acest lucru poate părea o distincție arbitrară. Asta pur și simplu pentru că este un proces complex, reprezentările necesită timp pentru a fi construite și este nevoie să se construiască pe straturile anterioare de înțeles. Uneori este nevoie de continuitate. Acesta este motivul pentru care LM nu vor fi înlocui pe deplin grafurile de cunoștințe, ci mai degrabă le vor stoca în memoriile lor interne sau distribuite.

CAPITOLUL 3

IAS PENTRU EXTRAGEREA CUNOȘTINȚELOR

3.1 ENTITĂȚI ȘI GRAFURI DE CUNOȘTINȚE

3.1.1 Fundal

Ideea dizambiguizării entităților (DE sau Named Entity Linking – NEL) a fost introdusă în 2006 de Răzvan Bunescu și Marius Pasca [BP06], care au adăugat cerința de a lega entitățile de un GC precum DBpedia [LIJ + 15]. Se știe că sistemele de dezambiguizare bune sunt create prin mai multe clase de algoritmi: (i) dezambiguizarea grafurilor prin detectarea comunității cu algoritmi precum Louvain [BGLL08] (de exemplu, Babelfy [MRN14a] și Recognyze [WKB18]); (ii) modele de limbaj statistic (de exemplu, DBpedia Spotlight [DJHM13]); sau (iii) modele neuronale (de exemplu, modelele propuse de Adel [AS19]).

Restul secțiunii descrie algoritmi personalizați pe baza arhitecturii pachetului Recognyze, precum și unele aspecte legate de evaluarea acestor algoritmi. Recognyze a fost dezvoltat în comun de cercetători de la Modul Technology, Universitatea Elvețiană de Științe Aplicate din Grisons și webLyzard de o echipă compusă din Albert Weichselbraun, Philipp Kuntschik și Adrian Brașoveanu. Mai multe detalii despre această arhitectură găsiți în [WKB18].

3.1.2 Varianța entităților

Entitățile pot să joace un rol în sarcini variate precum extragerea cunoștințelor (KE) sau extragerea relațiilor (ER), sarcini care pot fi definite pe baza tratamentului mențiunilor (de exemplu, șirul extras din text) sau a legăturilor (link-urilor) [WBKN19]:

- NER – Named Entity Recognition (Recunoașterea entităților) – se cere identificarea formei de suprafață a entității, poziția și tipul ei.
- NEL – Named Entity Linking (Dizambiguizarea entităților) – se cere în plus și legarea entităților la un GC precum DBpedia.
- SF – Slot Filling (Umplerea sloturilor sau Extragerea Relațiilor) – se cere completarea proprietăților lipsă pentru entitățile descoperite.
- (O)KE – (Open) Knowledge Extraction (Extragerea liberă a cunoștințelor) – se cere descoperirea tuturor entităților și relațiilor dintre ele dintr-un text liber.

Figura 3.1 prezintă legăturile dintre aceste sarcini. Un text DBpedia de pe pagina dedicată lui Edward Thorp este prezentat împreună cu entitățile incluse. Toate sistemele de dizambiguizare a

entităților vor trebui să colecteze mențiunile entităților și legăturile lor cu un GC țintă. Dacă sistemele oferă și detalii despre fiecare entitate, ar putea fi capabile să concureze și în provocări de completare a sloturilor. Un sistem DE este numit în general un adnotator automat, prin urmare vom folosi adesea acești doi termeni (sistem și adnotator) interschimbabil. A dezambigua între om și mașină adnotatoare (automate), vom folosi tipul (de exemplu, uman sau automat). Daca nu este atribuit tipul, se va presupune că adnotatorul este o mașină.

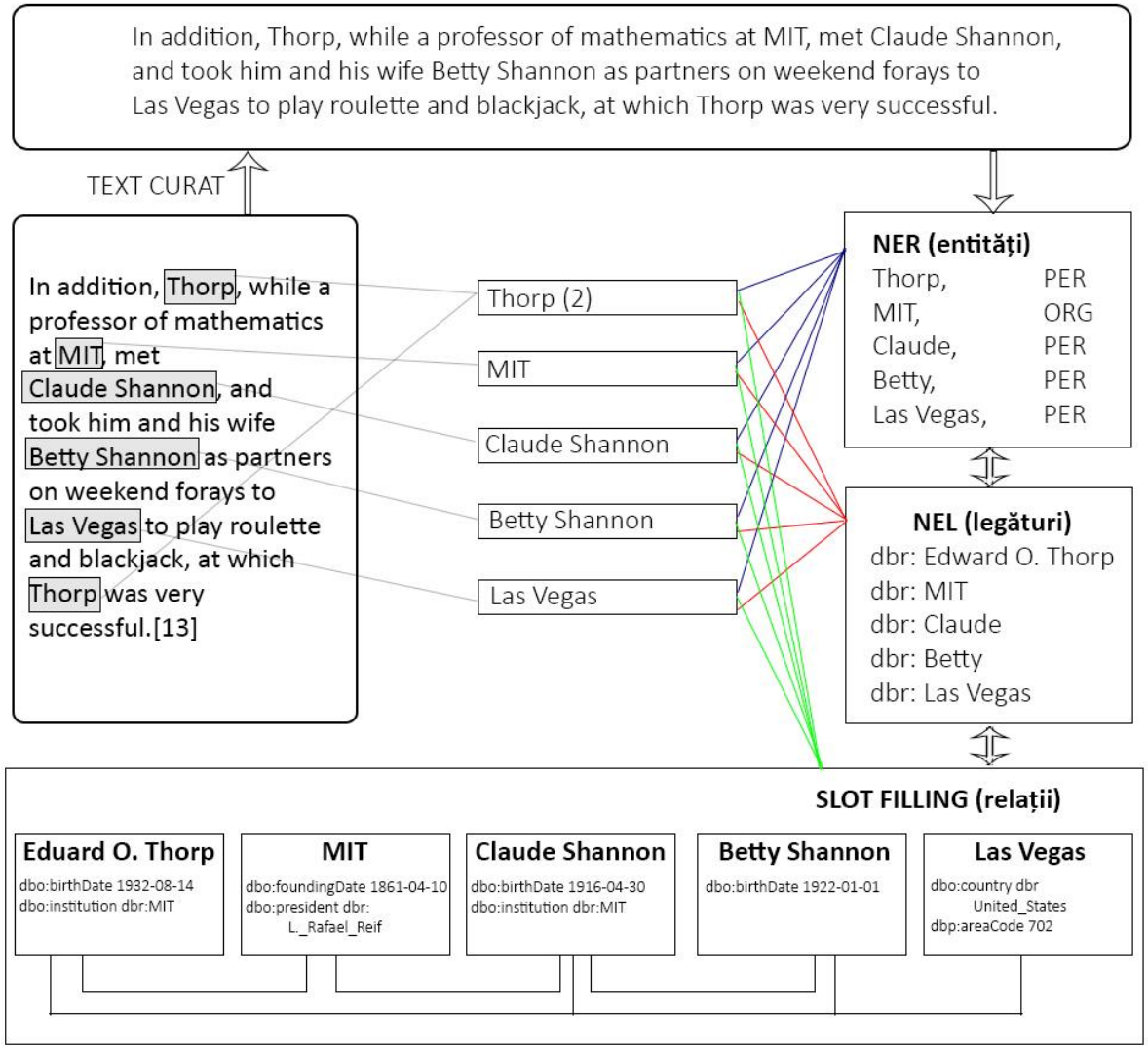

Figura 3.1 Legăturile dintre extragerea entităților (NER), dizambiguizare (NEL) și completarea sloturilor (extragerea relațiilor).

Definiție. Un sistem de dizambiguizare a entităților leagă o mențiune M a formei de suprafață a entității Si din documentul D de entitatea țintă Ei din GC. O pereche (xi,yi) reprezintă poziția mențiunii în interiorul unui document.

O mențiune este un șir de caractere care conține numele entității, deseori referit ca și formă de suprafață.

Adnotările standard de aur create de oameni trebuie să respecte ghidurile de adnotare oficiale. Astfel de linii directoare respectă adesea regulile exprimate în publicațiile anterioare precum cele din CoNLL 2003 [SM03]. Deoarece o anumită variație este așteptată între diferite limbi sau adnotatori, un judecător este obligat să soluționeze conflictele.

Problema entităților imbricate (de exemplu, NY Knicks poate fi legată atât de New York cât și echipa NY Knicks), precum și cea a variației numelui (de exemplu, numele diferite sub care o entitate poate fi prezentă într-un text) apar destul de frecvent în dizambiguizări. Întrebarea principală ne preocupă pentru restul secțiunii este: Care este impactul variației de nume asupra sistemelor de dizambiguizare a entităților?

Definiție. Varianța de nume înseamnă multitudinea de nume, titluri sau abrevieri atribuite unei singure entități dintr-un document.

Este o problemă cheie pentru dizambiguizare. Pentru a ne asigura că sistemele noastre sunt competitive ar fi important să dezambiguizăm diferitele mențiuni diferite către aceeași entitate. Îmbunătățirea gestionării variației numelui este, de asemenea, echivalentă cu îmbunătățirea acoperirii și ar trebui să conducă, de asemenea, la îmbunătățirea metricii de recall. *Prințul Charles*, de exemplu, poate fi numit atât *Prinț de Wales*, cât și *Duce de Edinburgh* astăzi, al doilea titlu fiind moștenit după moartea tatălui său, dar în viitor, la fel de bine poate fi numit *Regele Charles* dacă moștenește tronul.

Problema este agravată deoarece diferite tipuri de entități pot avea variațe suplimentare. Numele persoanelor pot include titluri, de exemplu, iar numele organizației pot include prefixe sau sufixe specifice fiecărei țări.

3.1.3 Varianța numelor și dizambiguizarea entităților

Următoarele pagini se bazează pe [WKB19] și discută trei metode care soluționează această problemă: (i) extinderea acoperirii numelor prin combinarea datelor din mai multe surse (de exemplu, seturi de date sau GC); și (ii) prin euristică; sau (iii) algoritmi ML care calculează automat aceste variante de nume.

Poate că cea mai firească idee a fost să colectăm pur și simplu variațele de nume din GC precum DBpedia și Wikidata și să le adăugăm la lexiconul original. Am testat atât impactul aceastei strategii atunci când colectăm variante dintr-un singur GC, cât și din multiple GC prin intermediul SPARQL Federation. Această metodă este în esență echivalentă cu un atac de lexicon pe problema variației de nume.

A doua idee a fost să împărțim numele candidaților în șiruri de caractere și apoi să le recombinăm pe baza unor reguli simple. O strategie simplă a fost de a recombina numai șirurile de caractere ale șirului original. O strategie mai complexă implică:

- Înlocuirea întregului șir cu sinonime.
- Folosirea de euristici care conțin expresii regulate, de exemplu sufixe și prefixe pentru organizații, pentru identificarea candidaților posibili, iar apoi modificarea și înlocuirea șirurilor corepunzătoare.

  Această metodă ar fi echivalentul unei euristici simple.

A treia idee implică construirea analizatorilor de nume, o extindere a generării algoritmice de nume bazată pe ideea de entropie. În Informatică entropia reprezintă zgomotul colectat de un semnal sau aplicație. Scorul de entropie evaluează câte nume de entități valide pot fi calculate din șirurile disponibile. Șirurile despre care se știe că sunt incluse în numele entităților (de exemplu, sufixe sau prefixe pentru organizații - cum ar fi Corp sau GmbH) pot fi premiate cu entropie mai mare.

Entropia care corespunde variației de nume {ti} a unei entități compuse din n tokens {t1, t2, ... tn} poate fi calculată cu formula [WKB19]:

$$H(\{t_i\}) = f_{\text{constr}}(\{t_i\}) \cdot \left[H_{\text{case}}(\{t_i\}) + H_{\text{classes}}(\{t_i\}) + \sum_{t_j \in \{t_i\}} H_{\text{token}}(t_j) \right] \quad (3.1)$$

În plus, H_{case} înlătură senzitivatea la majuscule, iar factorul f_{constr} înlătură cazurile care pot conduce la probleme sintactice.

O implementare alternativă a acestei metode a folosit versiunea Java a libSVM. Această metodă a folosit un set divers de caracteristici, inclusiv, dar nu limitat la: (i) morfologice (de exemplu, sensibilitate la majuscule, punctuație); (ii) sintactice (de exemplu, prepoziții, pronume); și (iii) caracteristici semantice (de exemplu, numărul de cuvinte care fac referire la locații, prenume sau nume comune, termeni de dicționar din mai multe limbi). Au fost obținute rezultatele optime după validarea încrucișată și căutarea în grilă a nucleului funcției bazei radiale ($C = 8$, $\gamma = 2–5$).

Putem considera această metodă ca fiind echivalentă cu un atac cu forță brută, dar implementarea sa poate fi realizată în mai multe moduri, după cum s-a explicat deja.

Evaluările au fost efectuate pe două seturi de date: N3 Reuters128 [RUH + 14], cunoscut ca fiind unul dintre cele mai dificile seturi de date pentru dizambiguizări datorită vechimii [AOOV20] și OKE215 [NGP + 15]. Cel mai bun rezultat a fost apoi comparat cu trei soluții concurente: AIDA [HYB + 11], Babelfy [MRN14a] și DBpedia Spotlight [DJHM13].

Tabelul 3.1 Varianța numelui și performanța pe Reuters 128 cu Recognyze. GN = Generare de nume. Baza = linie de bază. Diferențele semnificative reprezentate cu bold.

Setare	LOC			ORG			PER			Toare		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
baza	63	54	58	72	34	46	57	23	33	66	39	49
a) proprietati	63	54	58	71	33	45	57	23	33	66	38	49
b1) Wikidata	14	41	20	40	**41**	40	12	**38**	19	21	41	28
b2) Wikipedia	61	54	57	69	33	45	58	25	35	64	39	48
b3) GeoNames	60	54	57	71	33	45	57	23	33	64	38	48
b4) baza+(b1,b2,b3)	14	41	21	39	**41**	40	12	**38**	19	21	41	28
c) GN algoritmi	54	**72**	62	35	**53**	42	68	**49**	**57**	43	**58**	50
d1) GN Wikidata	52	54	53	71	38	50	59	26	36	61	42	50
d2) GN Wikipedia	58	52	55	68	35	46	60	29	39	63	39	48
d3) GN GeoNames	48	53	51	70	33	45	57	23	33	58	38	46
d4) baza+(d1,d2,d3)	46	53	50	70	38	50	61	30	40	58	42	49
e1) analizor nume (euristic)	64	52	57	47	**44**	46	60	**56**	**58**	54	**48**	51
e2) analizor nume (SVM)	65	51	57	33	**47**	39	55	**47**	**50**	42	**48**	45
f) baza+(a,c,d1,e1)	53	**70**	61	61	**52**	**57**	60	**56**	**58**	58	**58**	**58**

Tabelul 3.1 prezintă rezultatele. Am folosit valorile clasice din evaluări: precizie, rechemare (recall) și F1 pentru cele trei tipuri de entități clasice (Persoane, Organizații, Locații).

Linia de bază (baza) nu a inclus niciun tratament pentru varianța numelui. Adăugarea proprietăților RDF (cazul a) nu a condus la îmbunătățiri. În mod similar, folosind oricare unul dintre GC singur nu a condus la îmbunătățirile așteptate, ceea ce sugerează că adăugarea multor nume duce pur și simplu la mult zgomot dacă nu este echilibrată printr-o metodă de calculare a denumirilor semnificative (de exemplu, analizoare de nume). Generarea algoritmică a numelui (cazul c) a condus la îmbunătățiri ale rechemării, exact cum am teoretizat, dar precizia a scăzut. Cu toate acestea, aplicând același algoritm la intrările suplimentare din mai multe GC (cazurile d1-d4) nu a condus la îmbunătățiri

semnificative. Analizoarele de nume (cazurile e1-e2) îmbunătățesc rechemarea și F1 pentru oameni (PER) și locații (LOC). Interesant este că îmbunătățirile pentru numele organizației pot fi observate și în versiunea combinată (baza + (a, c, d1, e1)). Interesant este că, în comparație cu celelalte sisteme, Recognyze a obținut cele mai bune rezultate în evaluările respective și a menținut întotdeauna o margine confortabilă când vine vorba de rechemare.

Tabelul 3.2 Varianța numelor și performanța sistemelor de dizambiguizare.

Corpus	Sistem	LOC			ORG			PER			Toate		
		P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Reuters 128	AIDA	44	64	52	76	29	42	50	49	50	53	43	47
	BabelNet	29	31	30	47	16	24	21	29	24	32	22	26
	Recognyze	**53**	**70**	**61**	61	52	57	60	56	58	58	58	58
	Spotlight	41	**70**	52	64	42	51	47	22	30	50	49	49
OKE 2015	AIDA	25	37	30	69	43	53	66	41	50	50	41	45
	BabelNet	21	35	26	67	40	50	55	14	22	40	26	32
	Recognyze	**62**	**73**	**67**	70	51	59	85	57	68	73	59	65
	Spotlight	50	72	59	81	50	62	56	11	18	61	36	45

3.1.4 Varianța numelor și lentilele

Îmbunătățirea variantei numelui este doar jumătate din poveste. Instrumentele de benchmarking este posibil să nu fie pregătite pentru astfel de îmbunătățiri și parțial scorurile obținute au fost oarecum mai mici decât așteptările (deși nu cu mult) tocmai datorită acestui aspect. Pentru a îmbunătăți această gestionare a instrumentelor de benchmarking, o soluție pentru gestionarea potrivirilor parțiale și a entităților imbricate a fost necesară. Restul subsecțiunii descrie o astfel de metodă publicată în [BWN20].

Ideea lentilelor sau obiectivelor a apărut datorită unei metode folosite de fotografi pentru a produce diferite tipuri de imagini: schimbarea obiectivelor. În esență, putem crea adnotări bazate pe o anumită specificație. Un obiectiv ar putea selecta cu lăcomie cel mai lung șir de potriviri pentru fiecare entitate, în timp ce altul ar putea selecta cel mai scurt, de exemplu. Un altul ar putea face ceva la mijloc, cum ar fi luarea în considerare a suprapunerilor entității. Se pot crea și tipuri speciale de lentile prin extragerea subseturilor de adnotări pe baza tastării (de exemplu, seturi de locații) sau GC (de exemplu, lentile DBpedia sau Wikidata). Acest lucru ne conduce la definirea lentilelor:

Definiție. O lentilă este o adnotare în care se utilizează o singură regulă pentru adnotarea unei anumite proprietăți, cum ar fi tipul, lungimea sau legătura pe parcursul întregului set de date.

Ne putem gândi la lentile ca fiind fie simplificări extreme ale adnotării liniilor directoare (de exemplu, în loc să creăm zece reguli pentru adnotarea șirurilor lungi, vom folosi o singură regulă) sau cazuri speciale (de exemplu, decidem să evaluăm numai tipurile sau link-uri Wikidata). Aceasta înseamnă că prin utilizarea mai multor obiective putem simula efectul diferitelor alegeri de proiectare asupra rezultatelor.

O încercare timpurie de a prezenta conceptul de corpus cu lentile este In Media Res [BWN20]. Toate documentele au fost colectate din surse publice (de ex., Wikipedia, Wikinews etc.) și puse la dispoziția publicului prin GitHub.

Tabelul 3.3 Exemple de ieșiri pentru lentile.

Exemplu	ΦMIN	ΦMAX	OMAX
Sir Patrick Stewart OBE	1: Sir Patrick Stewart OBE	1: Sir 2: Patrick Stewart 3: OBE	1: Sir Patrick Stewart OBE 2: Sir 3: OBE
MLB Advanced Media (MLBAM)	1: MLB Advanced Media (MLBAM)	1:MLB Advanced Media 2: MLBAM	1: MLB Advanced Media (MLBAM) 2: MLBAM
Burbank, California	1: Burbank, California	1: Burbank 2: California	1: Burbank, California 2: California
Seinfeld	1: Seinfeld	1: Seinfeld	1: Seinfeld

Acest corpus explorează convențiile de denumire pentru entitățile care apar frecvent în mass-media precum francizele sau emisiunile TV.

Pentru a simplifica procedura de adnotare, următoarele tipuri de lentile au fost dezvoltate:

- ΦMIN – număr minin de entități.
- ΦMAX – număr maxim de entități fără potriviri parțiale.
- OMAX – număr maxim de entități cu potriviri parțiale.

Tabelul 3.3 oferă câteva exemple pentru aceste reguli de adnotare. Am inclus câte un exemplu pentru fiecare tip de entitate (Persoană, Organizație, Locație), dar și un exemplu pentru o Operă (Work).

O evaluare realizată pe partiția primară a datasetului In Media Res e prezentată în tabelul 3.4. Două dintre sistemele examinate (Recognyze și DBpedia Spotlight) par să beneficieze mai mult de folosirea acestor lentile, dar toate sistemele au arătat îmbunătățiri de până la 4 procente.

Tabelul 3.4 Performanța sistemelor pe o partiție cu lentile.

Corpus	*Sistem*	mP	mR	$mF1$	MP	MR	$MF1$
Core set ØMIN	AIDA	0.47	0.48	0.47	0.43	0.48	0.43
	Spotlight	0.53	0.43	0.48	0.35	0.42	0.37
	Recognyze	**0.61**	**0.52**	**0.56**	**0.52**	**0.50**	**0.51**
Core set ØMAX	AIDA	0.49	0.48	0.49	0.45	0.48	0.44
	Spotlight	0.55	0.43	0.48	0.35	0.40	0.36
	Recognyze	**0.62**	**0.54**	**0.58**	**0.55**	**0.52**	**0.53**
Core set OMAX	AIDA	0.49	0.48	0.49	0.45	0.48	0.44
	Spotlight	0.51	0.57	0.54	0.51	**0.58**	0.52
	Recognyze	**0.65**	**0.61**	**0.64**	**0.61**	0.57	**0.59**

3.1.5 Discuție

Au fost discutate mai multe strategii pentru implementarea variantei de nume în sistemele de dizambiguizare a entităților: (i) extragerea variantelor de nume din GC; (ii) generarea de nume prin algoritmi; și (iii) analizoare de nume. S-a găsit o combinație a acestor strategii care poate conduce la îmbunătățiri față de instrumentele concurente (de exemplu, AIDA, Babelfy), dar doar o margine mai redusă comparativ cu DBpedia Spotlight.

S-a descoperit că pentru a aprecia pe deplin impactul variantei de nume asupra dizambiguizării, trebuie făcute o serie de modificări la procedurile de evaluare. Ideea de stiluri de adnotare reduse care se concentrează pe o singură proprietate (de exemplu, mențiune, tip, link) numit lentile a fost introdusă. Dezvoltarea lentilelor este doar o posibilitate care poate fi luată în considerare pentru a evalua pe deplin impactul variantei de nume. Un corpus a fost dezvoltat pentru a testa această ipoteză. Sistemele examinate au oferit într-adevăr rezultate mai bine cu aceste noi setări de evaluare, care au considerat multiple posibilități.

3.2 SENTIMENT ȘI EMOȚIE

3.2.1 Fundal

Analiza sentimentelor (SA) este considerată o tehnologie umbrelă [CLH11]. Aceasta este multistrat, deoarece trebuie să combine sintaxa (de exemplu, etichetarea POS), semantica (de exemplu, NER, dezambiguizarea sensului cuvântului) și pragmatica (de exemplu, polaritate, aspect, sarcasm).

Problema de bază a SA este identificarea afirmațiilor sunt pozitive, negative, neutre sau ambivalente față de un anumit obiect sau idee. Analiza emoțiilor (EA) oferă o clasificare a emoțiilor cu granulație fină, pe măsură ce emoțiile sunt mapate pe categorii emoționale. SA ne oferă o evaluare a emoției, în timp ce EA încearcă să ne ofere o imagine mai bună legând emoția de un model mai complex.

3.2.2 Modele de categorizare afectivă specifice domeniului

Această secțiune se bazează pe [WSB + 21].

Un model de categorizare afectivă poate fi considerat ca o taxonomie pe care se bazează un model de clasificare afectivă. Conține etichetele care vor fi prezise de clasificatorul emoțional. Unele modele bine cunoscute includ Hourglass of Emotions al lui Cambria și versiunea revizuită a lui Susanto [SLCC20].

Modele afective specifice domeniului sunt create pentru a calcula anumiți indicatori. Dacă o organizație, de exemplu, își definește marca folosind anumite fraze sau cuvinte, poate dori să urmărească asocierile dintre aceste cuvinte cheie și numele său. Într-un astfel de caz, poate fi necesară definirea unui model specific acestui domeniu, deoarece un model definit pentru industria de divertisment nu va funcționa neapărat pentru industria media.

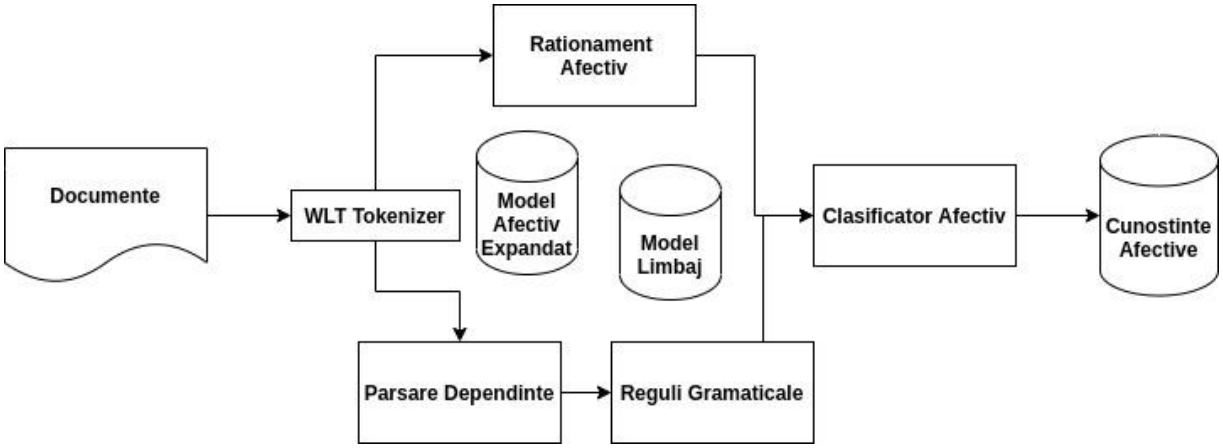

Figura 3.2 Arhitectura de clasificare afectivă.

Procesul începe prin adnotarea unui model existent (de exemplu, modelul vechi) cu entități din GC precum ConceptNet [SCH17] și Wikidata [Vra13] prin extragere de fraze și sinonime / antonime legate de termenii sămânță. Adnotările sunt apoi contextualizate prin extragerea de propoziții care conțin concepte GC. Exemplele de propoziții prezintă utilizarea unui termen în context. Apoi un algoritm de dezambiguizare a sensului cuvântului este aplicat pentru a calcula corespondența asociațiilor de categorii bazate pe sensurile termenului. Algoritmul se repetă pentru diferite sensuri și folosește un model de limbaj (LM) pentru a identifica sensurile corespondente. Un set de categorii rafinate este apoi calculat pe baza utilizării termenului. Pentru concepte care nu sunt incluse în GC, algoritmul folosește exemplele din corpus. Următorul pas este expansiunea afectivă, în care se adaugă și antonimele și sinonimele.

Procesul de extragere a cunoștințelor afective folosește apoi modelul extins și transformă sensurile care sunt utilizate de LM precum BERT în caracteristici. Un vector caracteristic este apoi

calculat folosind simbolul, propoziția și arborele de analiză a dependenței corespunzător (DP). O căutare de proximitate este realizată pentru a calcula valoarea categoriei afective în contextul evaluat. Scorul pentru categorie este apoi calculat luând în considerare negația și modificatorii bazați pe DP și reguli gramaticale.

Pentru această evaluare a fost creat un corpus special colectat din WikiNews. Regulile de adnotare s-au bazat pe provocări anterioare (de exemplu, SemEval [CNJA19], sau SMM4H [SBF + 18]), fiind furnizate exemple pentru fiecare dintre cele patru categorii cu valențele lor (introspecție, temperament, plăcere, nerăbdare). Pe lângă liniile directoare, adnotatorii au primit, de asemenea, tabelele care explicau categoriile emoționale actualizate din [SLCC20] și o listă de declanșatoare pentru contrariile polare pentru fiecare categorie afectivă. În cele din urmă, lista a fost inclusă în ghidul de adnotare. Declanșatorii conțineau liste de cuvinte care pot oferi indicii pentru categoriile emoționale. Mai multe exemple din publicații anterioare (de exemplu, [SLH + 18] și [PHM + 19]) au fost de asemenea selectate pentru echilibrare. Fiecărui participant i s-a cerut să adnoteze 120 de propoziții și să furnizeze indicii despre polaritate, categorii afective și emoție dominantă. Eticheta Unknown a fost utilizată pentru marcarea cazurilor în care o emoție dominantă a fost absentă, în timp ce eticheta None a fost atribuită afirmațiilor fără emoții. Corpusul a fost creat sub supravegherea expertului cu experiență relevantă în analiza sentimentelor și extragerea de entități. Adnotatorii au avut ocazia să consulte expertul în timpul creării adnotărilor pentru cazuri dificile. Emoția dominantă a fost selectată prin media scorurilor pentru categoriile afective și polaritate.

Categoriile actualizate (de exemplu, temperamentul, introspecția, atitudinea și sensibilitatea) din Hourglass of Emotion [SLCC20] și concepte asociate din aceeași publicație au fost utilizate pentru a evalua SenticNet 5, deoarece a șasea versiune nu era încă disponibilă public la momentul respectiv [CLX + 20]. Top 20 de termeni, cu excepția bolilor au fost selectați, în timp ce noii termeni au fost adăugați manual pentru a crea un set de semințe echilibrat.

Tabelul 3.5 Rechemare pentru emoția dominantă.

categorie	BERT	DistilBERT	BERT+GR	DistilBERT+GR
T+ calmness	0.62	0.46	**0.75**	0.68
T- anger	0.55	**0.65**	0.45	**0.65**
I+ joy	0.37	**0.46**	0.40	0.43
I- sadness	0.76	0.80	0.74	**0.83**
A+ pleasantness	0.62	0.64	0.65	**0.67**
A- disgust	0.68	**0.69**	**0.69**	0.68
S+ eagerness	0.38	0.36	**0.46**	0.36
S- fear	**0.90**	0.87	0.80	0.70
overall	0.61	0.63	0.62	**0.64**

Ceea ce este interesant de observat este că multe propoziții au avut valoare neutră. Acest lucru se datorează parțial faptului că afirmațiile au fost colectate din surse jurnalistice (Wikinews), deci trebuiau să fie cel puțin în teorie imparțiale. O altă observație este faptul că deseori a fost nevoie de un singur declanșator pentru a avea un impact asupra propozițiilor non-neutre. Scorurile sunt sporite în continuare de adăugarea analizelor de dependență (DP) și a regulilor gramaticale (GR).

Tabelul 3.5 prezintă câștigurile de performanță obținute prin aplicarea Transformers precum BERT și DistilBERT. Au fost testate și alte câteva modele (de exemplu, RoBERTa, XLNet), dar de când am descoperit că modelul clasic BERT (bert-base-uncased) și modelul distilat (distillbert-base-uncased) au avut cele mai bune rezultate timpurii, le-am luat în considerare doar pe acestea în evaluările finale.

3.2.3 Discuție

Un bias negativ a fost confirmat în standardul aur, deoarece mai multe propoziții negative au fost găsite. Acest lucru este confirmat de literatură, deoarece se știe că articolele politice tind să aibă conotații negative [LEB12].

Corpusul și evaluarea asociată demonstrează că metoda propusă funcționează bine cu LM-uri de diferite dimensiuni, chiar și în medii în care resursele sunt rare. Evaluarea arată, de asemenea, că metoda poate fi utilizată pentru a actualiza abordările anterioare (de exemplu, pentru a include negarea, dacă este necesar).

3.3 VERIFICAREA FAPTELOR

3.3.1 Fundal

Această secțiune se bazează pe [BA19] și versiunea sa extinsă publicată în [BA20a].

Există mai multe cazuri când este necesară verificarea faptelor. Dacă o dată de naștere prezentă într-un GC este greșită, putem verifica în mai multe surse. Când nu există nicio dovadă clară a ceea ce s-a întâmplat, este posibil să fie necesară verificarea faptelor.

În mediile online, aceasta corespunde urmăririi provenienței datelor pentru a înțelege cine a lansat o știre și de ce. O altă posibilitate este să utilizăm textele așa cum au fost scrise împreună cu unele date de context (dacă sunt disponibile). Detectarea știrilor false ar fi în general o astfel de instanță. Știri false pot fi văzute ca parte a altor câteva clase mai mari de probleme, inclusiv detectarea propagandei și verificarea faptelor [TVCM18]. Pentru restul secțiunii o vom considera ca făcând parte din clasa de verificare a faptelor.

Detectarea propagandei este o clasă mult mai mare care depășește simpla verificare, deoarece poate include și imagini sau videoclipuri de natură înșelătoare și de aceea viziunea este importantă atunci când o analizăm. Un sondaj privind automatizarea identificării propagandei poate fi găsit în [MCB + 20]. Știrile false sunt acum un domeniu interdisciplinar larg, greu de caracterizat pe deplin doar prin articole NLP. Diverse perspective pot fi găsite într-un set recent de sondaje [PCLG21].

3.3.2 Știri false

Definiția verificării faptelor sugerează că nucleul sarcinii este o problemă de predicție:

Definiție (bazată pe [PZS + 20]). Având două secvențe, un set de afirmații S = {s1 ..., sn} și setul de fapte cunoscute F = {f1, .., fn}, se cere să prezicem relația dintre cele două secvențe.

Relația este gradul de sprijin și poate fi modelată ca o etichetă care indică dacă afirmațiile sunt susținute sau infirmate de faptele cunoscute. Faptele cunoscute pot fi restul atributelor cunoscute din setul de date sau caracteristici calculate. Majoritatea definițiilor pentru știrile false, deoarece au fost definite prin lentilele politicii, nu consideră că predicția relației este nucleul problemei.

Definiția clasică a știrilor false se bazează pe studiul alegerilor din 2016 din SUA [AG17]:

Definiție. Știrile sau știrile parțiale pot fi considerate false în cazul în care conținutul lor este dovedit fals.

Problema ML corespunzătoare este clasificarea, fie binară, fie multi-clasă, în funcție de setul de date. Intrarea conține instrucțiuni scurte. De asemenea, este posibil să furnizăm informații suplimentare (de exemplu, data, locația). Ieșirea este o etichetă (de exemplu, binară sau cu granulație fină).

3.3.3 Semantica știrilor false

Pentru a oferi capacitățile semantice necesare pentru identificarea știrilor false, conducta utilizată pentru această sarcină include:

- Adnotări - conducta de bază include etichetarea POS, entități și sentiment;
- Relații - calculate din text și GC;
- Modele neuronale - modele DL care includ încorporări.

Relațiile au fost calculate din text și GC. Două tipuri de relații au fost extrase din text i) între entități (de exemplu, verbe între substantive proprii), sau ii) între entități și obiecte (de exemplu, verb între subiect și obiect).

Relațiile între entități au fost, de asemenea, extrase din DBpedia, acolo unde a fost posibil.

Seturile de date Liar [Wan17] și Politifact [RCJ + 17] au fost publicate în 2017 și conțin date similare extrase de pe situl politic de verificare a afirmațiilor Politifact. Principala cerință pentru ambele seturi de date este adnotarea unor texte scurte cu șase clase dezechilibrate care reprezintă gradele lor de adevăr (de la Adevărat la Pants on fire). Spre deosebire de tweet-uri, aceste afirmații nu conțin limbajul specific prescurtat al textelor de pe Internet (de exemplu, emojis, retweets), dar mai degrabă un limbaj natural.

Bogăția de date disponibile pentru aceste seturi de date permite mai multe configurații experimentale: (i) numai propozițiile (în esență textele, etichetat cu T); (ii) atributele originale sau T + A (de exemplu, declarații plus alte caracteristici prezente în seturile de date); (iii) text plus caracteristici semantice etichetate T + R; și (iv) toate caracteristicile combinate (sau TOATE). Trebuie remarcat faptul că Politifact conține doar texte. Acest lucru sugerează că T + A nu este necesar în realitate.

Modelele DL folosesc în special modelul Glove 300 încărcat cu Keras API. Embbeddings sunt plasate pe primul strat după intrări, deoarece plasarea respectivă a fost considerată bună pentru procesarea seturilor de date mici [QSF + 18].

Tabelul 3.6 Acuratețea pe setul de testare de la Liar.

Model	T	T+R	TOATE
ML Clasic			
Multinomial Naive Bayes	0.224	0.244	**0.262**
SGDClassifier	0.239	0.235	**0.255**
Logistic Regression (OneVsRest)	0.240	0.260	**0.273**
Random Forest	**0.215**	0.215	0.212
Decision Trees	0.226	0.249	**0.262**
SVM	0.255	0.275	**0.294**
Deep Learning			
CNN	0.241	0.270	**0.289**
BasicLSTM	0.245	0.289	**0.326**
BiLSTM Attention	0.419	0.448	**0.499**
GRU Attention	0.450	0.496	**0.539**
CapsNet	0.565	0.598	**0.649**

Tabelele 3.6 și 3.7 prezintă modelele considerate și scorurile acestora și raportează scorurile de precizie stabilite. Am împărțit tabelele în ML clasic (de exemplu, modele statistice sau în general modele pre-DL) și Deep Learning. Modelele DL obțin rezultate mai bune. Aceasta nu este întotdeauna o concluzie predeterminată, deoarece este știut că ansamblurile bazate pe Conditional Random Forests (CRF) funcționează bine pentru probleme semantice [PRT16].

Pentru modelele statistice [HTF09], îmbunătățirile obținute cu caracteristici semantice suplimentare sunt mici (de exemplu, 2-3%). Scorurile sunt mici și doi clasificatorii arată aceleași rezultate pentru Politifact (de exemplu, arbori de decizie și regresia logistică). Interesant este faptul că

Random Forest este singurul clasificator care nu prezintă îmbunătățiri cu caracteristici semantice suplimentare, dar s-a folosit modelul de bază și nu un ansamblu așa cum este sugerat în mod tipic în literatură. Supremația SVM este confirmată pentru ambele seturi de date pentru această clasă de algoritmi.

Tabelul 3.7 Acuratețea pe setul de testare de la Politifact.

Model	T	T+R	TOATE
ML Clasic			
Multinomial Naive Bayes	0.263	0.295	**0.296**
SGDClassifier	0.262	0.294	**0.295**
Logistic Regression (OneVsRest)	0.246	0.269	**0.269**
Random Forest	**0.244**	0.229	0.229
Decision Trees	0.246	0.269	**0.270**
SVM	0.262	0.281	**0.282**
Deep Learning			
CNN	0.203	0.231	**0.244**
BasicLSTM	0.245	0.287	**0.282**
BiLSTM Attention	0.371	0.422	**0.422**
GRU Attention	0.415	0.451	**0.452**
CapsNet	0.473	0.523	**0.524**

Modelele DL utilizează codarea la cald a etichetelor clasei. Folosesc TensorFlow [ABC + 16], Keras [Cho17], pierderea categorică a crossentropiei și optimizator Adam [KB14]. Pașii de preprocesare au inclus: curățarea documentului, extragerea de cuvinte cheie, tokenizare, transformarea textului în secvențe și căptușeală. Pe lângă CNN și LSTM, restul modelelor folosesc modelul Glove300. Modelele pre-antrenate utilizate cât mai mult posibil. Aceeași rată de învățare (LR) și dimensiune a lotului sunt folosite peste tot și aceeași condiție de oprire.

LSTM-ul de bază este un LSTM cu dimensiunea de 300, GlobalMaxPool, spatial dropout la 2 și layere ascunse dense cu activare softmax. Hiperparametri includ mărimea lotului de 256, 20 de epoci și LR=0.001.

BiLSTM [CN16] este bazat pe CuDNNLSTM cu atenție, dropout și recurring dropout setate la 0.25 și un set de straturi dense ascunse activate cu softmax. Aceeași parametri ca la modelul de bază sunt folosiți.

GRU [ITA+16] folosește setări și hiperparametri de la modelul anterior.

CapsNet este bazat pe [FK19, KJPC20]. Conține straturi capsulă ca și înlocuire pentru GlobalMaxPool. Un BidirectionalGRU cu dimensiunea de 128 ste activat de ReLU cu dropout și dropout recurent setat la 0.25. Rezultatul este trimis unei singure unități și trecut printr-o sigmoidă. Parametri

adiționali includ 10 capsule cu dimensiunea de 16 și 5 rutări. Numărul de epoci de antrenament a fost setat la 5.

3.3.4 Discuție

Modelele statistice nu sunt bine pregătite pentru această sarcină. Experimentele sugerează că semantica și o preprocesare bună pot fi cheia unor rezultate mai bune, precum creșterile de precizie de până la 4,2% obținute prin adăugarea semanticii sau de până la 10% pentru modele cu atenție.

Aceste scoruri sunt menite să fie interpretate ca linii de bază și nu ca și cele mai bune scoruri posibile pentru această sarcină. Calculele de bază sunt centrale pentru dezvoltarea rapidă, mai ales dacă codul poate fi utilizat în diferite setări (de exemplu, cercetare sau producție). Este o distincție importantă, deoarece înseamnă că, folosind aceste tehnici, ar trebui să fie posibil să se construiască rapid un bun clasificator care să se bazeze pe modele DL cunoscute. Acesta este principalul motiv pentru care majoritatea modelelor au fost limitate la funcționalitatea lor de bază.

CAPITOLUL 4

EXPLICABILITATEA IAS

4.1 EXPLICABILITATEA DIZAMBIGUIZĂRII ENTITĂȚILOR

4.1.1 Introducere la evaluarea dizambiguizării entităților

Primele evaluări ale dizambiguizării entităților au furnizat pur și simplu metricile clasice (precizie, rechemare și F1) fără a adăuga nicio interpretare. Majoritatea evaluărilor nu au fost într-adevăr standardizate, dar în schimb s-au bazat pe API-urile de măsurare furnizate de librării ML precum Scikit-learn sau PyTorch. Astfel de API-uri au fost concepute pentru a livra rezultate în format caseta neagră, ceea ce înseamnă că utilizatorii au primit scorurile finale, dar fără prea multe detalii cu privire la ce a mers prost. Desigur, acest lucru a fost ideal pentru cazuri în care rezultatele au fost bune, dar mai puțin ideal pentru testele pline de erori.

Evaluările NEL implică de obicei mai multe componente:

- *Set de date (standard de aur)* – este setul de date pe care va fi rulată evaluarea [HLAN12].
- *GC (Graf de cunoștințe)* – este graful de referință folosit pentru adnotări. Din păcate un GC este schimbat destul de des și legături spre alte GC pot fi adăugate sau șterse [RUN18].
- *Adnotator* – este un sistem de dizambiguizare automată a entităților care va genera ieșirile [RUN18].
- *Cluster NIL* – este o componentă prin care un sistem de dizambiguizare automată va grupa entitățile, de exemplu o persoană care e identificată prin mai multe nume ar trebui să fie legată la o singură intrare dintr-un graf de cunoștințe [HNR14].
- *Marcator (Scorer)* – o componentă care va calcula scorurile finale. Va returna metricile de performanță și dacă este posibil niște explicații de bază ale acestora.

Fiecare componentă poate genera alte tipuri de erori și din această cauză depanarea IAS este dificilă.

La evaluarea rezultatelor se raportează metricile clasice, cum ar fi precizia (P), rechemarea (R) și scorul F1 (F1). Uneori, scorul poate fi influențat și de suprapuneri sau potriviri parțiale. În general este de presupus că entitățile se pot găsi în una dintre următoarele situații: i) potriviri perfecte (de exemplu, dacă textul menționii se potrivește cu întregul șir al formei de suprafață); ii) complet conținute în formele suprafeței de aur (de exemplu, dacă entitatea returnată este conținută într-un exemplu furnizat în setul de aur); sau (iii) suprapuse parțial cu setul de aur acceptat (de exemplu, entitatea ar putea extinde mai mult decât ceea ce a fost înregistrat în setul de aur).

Câteva suite de benchmarking pentru dizambiguizarea entităților includ BAT2 [CFC13], neleval [HNR14] și Gerbil [RUN18]. Majoritatea acestor sisteme sunt construite în jurul filozofiei cutiei negre și oferă în general doar rezultatele evaluării, dar mai puține explicații sau vizualizări.

4.1.2 O taxonomie a erorilor pentru evaluarea dizambiguizării entităților

După ce am analizat ieșirile suitei neleval [HNR14] pentru mai mulți adnotatori, inclusiv DBpedia Spotlight [DJHM13], Babelfy [MRN14b], AIDA [HYB + 11] și Recognyze [WKB18], am încercat să creăm o clasificare a erorilor. Am început prin colectarea diferitelor erori observate în documente din seturile de date KORE50 [HSN + 12], Reuters128 [RUH + 14] și RBB150 [BNWS16]. După o discuție preliminară, am decis categoriile din taxonomie. Apoi am procedat la adnotarea a 50 de documente cu neleval și am etichetat manual erorile pe baza taxonomiei. Un rezumat al observațiilor preliminare poate fi găsit în Tabelul 4.1.

Tabelul 4.1 Exemple de erori

Sistem (s)			**Standard Aur (g)**		**Eroare**	
$Link_s$	**ET_s**	**Forma**	**$Link_g$**	**ET_g**	**Tip**	**Cauza**
Bruce_Willis	ORG	expiration	-	-	GC	Redirects
(de.)2009	LOC	2009	-	-	GC	Wrong Type
United_States	LOC	U.S.	-	-	DS	Missing Annotation
New_York_City	LOC	New York	New_York	LOC	DS	Wrong Annotation
(de.)Berlin	LOC	Berlin	Berlin	LOC	DS	Different Language
JFK	PER	Kennedy	JPK	PER	AN	Same-Type
Beck	ORG	Beck	Jeff_Beck	PER	AN	Cross-Type
Barack_Obama	PER	Malia Obama	NIL	PER	NIL	Wrong Cluster
NIL	ORG	Knicks	New_York_Knicks	ORG	NIL	Partial Match
Miles_Davis	PER	Davis	Miles_davis	PER	SE	Correct Redirect

Au fost găsite cinci clase de erori, chiar dacă una dintre aceste clase este mai puțin frecventă.

Erorile GC sunt, în general, erori care au avut originea în GC. Probabil au fost colectate din arhive sau versiuni live vechi. Exemplele includ adnotări în diferite limbi (de exemplu, adnotarea în germană în loc de engleză), forme de suprafață greșite (de exemplu, cuvinte care nu au nimic în comun cu ținta) sau redirecționări. Uneori, astfel de erori pot fi identificate numai când cercetătorii au acces la versiunea CG utilizată pentru adnotările standardului de aur.

Erorile DS sunt probabil cele mai problematice, deoarece sunt mai greu de remediat dacă corpusurile nu sunt publicate cu o licență permisivă. Unele dintre aceste erori pot fi similare cu erorile GC (de exemplu, o formă de suprafață greșită, o adnotare în limbaj diferit), dar au fost observate în DS. Astfel de erori pot fi uneori fixate prin utilizarea lentilelor, așa cum s-a discutat în capitolul 3.1.4.

Erorile de adnotare (AN) sunt cele mai frecvente erori observate. Ele pot include de la diferite tipuri, până la abrevieri greșite sau termeni generici returnați în locul unei entități. Aceste erori sunt cauzate de setările adnotatorului și pot fi uneori eliminate dacă se utilizează alte setări.

Erorile de grupare NIL sunt răspândite între potrivirile parțiale sau partajarea numelor între clustere (de exemplu, atunci când rolurile sau titlurile sunt alocate mai multor clustere). Algoritmii timpurii de clusterizare au folosit rareori mecanisme de rezoluție a co-referinței și au fost mai predispuși la erori [Rad15]. Co-referințele sunt deosebit de necesare pentru a detecta cazurile în care este utilizat un proxy în locul numelor entității.

Cele mai puțin frecvente erori au fost de marcare (sau scoring – SE). Am descoperit doar o astfel de eroare care a fost cauzată de redirecționare (de exemplu, un link Miles Davis care a fost în schimb o redirecționare a legăturii corecte a fost clasificată ca legătura corectă).

Au fost, de asemenea, colectate diferite numărări de exemple pentru aceste erori. Codul utilizat pentru calculul erorilor a fost ulterior inclus în Orbis.

4.1.3 Orbis

Clasamentele sunt folosite în multe provocări ca mijloace pentru promovarea concurenței și motivarea echipelor de cercetare pentru a-și îmbunătăți rezultatele. Cu toate acestea, feedback-ul oferit dezvoltatorilor este limitat la rangul lor pe o tablă sau la rezultate și poate să nu fie suficient pentru identificarea mijloacelor de îmbunătățire a acestor sisteme. Următoarele pagini descriu Orbis, un sistem publicat în [OKBW18].

Întrucât generația actuală de instrumente de adnotare rareori publică setările cele mai bune și liniile directoare de adnotare sunt - după cunoștințele noastre – rareori disponibile în formate citibile de mașină, reproducerea exactă a rezultatelor nu este întotdeauna posibilă.

Orbis a fost conceput având în vedere principiile Open Data și, prin urmare, acceptă metodologia de publicare a datelor FAIR [JdMAJ + 20]. Acronimul înseamnă „Findable, Accesible, Interoperable și Reusable". Aceste principii au fost create pentru a sprijini reproductibilitatea cercetării și aproape toate sunt implementate în Orbis, cu excepția referințelor calificate la alte metadate și a metadatelor care pot fi căutate.

Orbis acceptă în prezent următoarele sarcini: (i) extragera entităților; (ii) dizambiguizarea entităților; (iii) extragerea relațiilor (ER sau umplerea sloturilor – US); și (iv) extragerea datelor de pe forumuri, care este un caz special de extragere a conținutului. [WBWO21]). Deși ultima sarcină nu este o sarcină axată pe entități, a fost considerată o sarcină NLP importantă, deoarece calitatea metadatelor

extrase este la fel de bună ca și calitatea textului extras. Orbis a fost dezvoltat prin multiple proiecte de cercetare. Un număr mare de seturi de date și adnotatori a fost integrat (câteva zeci).

Conducta Orbis a fost construită în jurul formatului NIF [HLAN12] pentru publicarea datelor NLP. A fost proiectat în jurul unei conducte centrale descrisă printr-un fișier de configurare YAML. Conducta centrală încarcă datele (de exemplu, standardele aurului, ieșiri adnotatoare) și o trimite către o componentă de evaluare care produce o matrice de confuzie și valorile de performanță. Rezultatele sunt convertite în diferite formate de ieșire și transformate în analize printr-un set de plugin-uri. Restul componentelor sunt plugin-uri concepute pentru a efectua o singură sarcină cum ar fi citirea standardelor de aur sau afișarea documentelor adnotate.

Interfața este construită în jurul unui panou dublu de rezultate din setul aur și adnotator. Utilizatorii pot alege între mai multe scheme de clasificare. Pentru fiecare schemă a fost implementat un alt tip de colorare pentru a anunța utilizatorii că trebuie să aștepte un comportament. Clasificatorii de bază sunt următorii:

- Entitate – fiecare entite este marcată cu o culoare diferită.
- Tip – fiecare tip de entitate este colorat diferit.
- Rezultat – culorile marchează tipul rezultatului (de exemplu, Pozitiv Adevărat, Pozitiv Fals, etc).

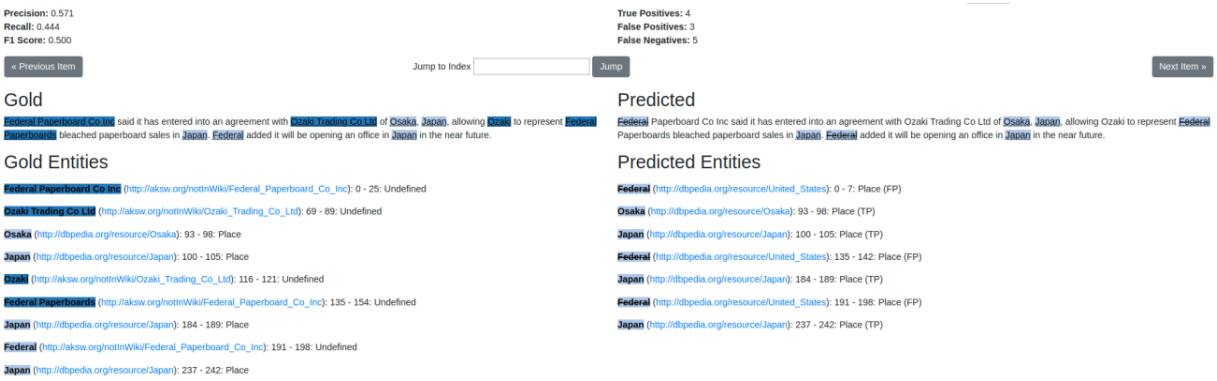

Figura 4.1 Clasificarea tipurilor în Orbis pentru dizambiguizarea entităților.

În funcție de sarcini, există și clasificări particulare și scheme de culoare asociate pentru ele. Versiunea curentă suportă:

Grupare – toate proprietățile care aparțin unei singure entități sunt colorate cu o singură culoare.

Paragraf – în loc de entități sau proprietăți, culorile identifică blocurile extrase corect (pentru sarcina de clasificare a textului extras din forumuri).

Erori – în loc de entități, clasele de erori sunt marcate cu culori diferite (doar pentru sarcini de clasificare automată a erorilor).

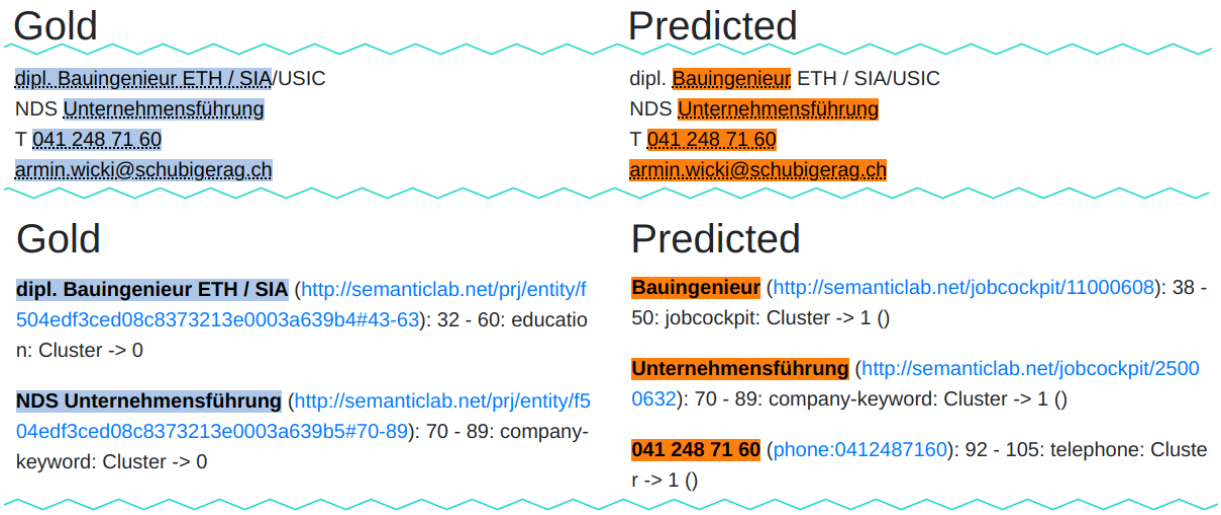

Figura 4.2 Clasificarea vizuală a relațiilor cu Orbis.

În partea de sus a tabloului de bord, Orbis include rezultatele evaluării arătînd tipul, metricile și rezultatele lor. Pentru depanare, un set redus al acestei funcționalități este afișat pe fiecare pagină.

4.1.4 Discuție

Această secțiune încheie discuția despre benchmarking explicabil și revizuiește câteva idei despre îmbunătățirea proceselor de benchmarking care au fost prezentate în [WBKN19].

După cum am văzut în secțiunile anterioare, depanarea proceselor de dizambiguizare este o meserie dificilă. Pot apărea probleme pe diferite straturi (de exemplu, GC, DS). In timp ce adnotatorii vor cauza majoritatea problemelor, marcatorii pot introduce și erori de notare.

Poate că cea mai gravă problemă este faptul că nu există niciun studiu legat de corectitudinea standardelor de aur și o metodă standard de rezolvare a problemelor acestora. În timp ce mai multe articole menționează această problemă (de exemplu, [vEMP + 16] și [JRN17]), nu sunt furnizate soluții. O altă problemă este aceea că adnotatorii pot fi optimizați pentru anumite seturi de date (de exemplu, prin antrenament excesiv pe acestea). Această problemă poate fi rezolvată dacă sunt publicate cele mai bune setări pentru fiecare adnotator și set de date, dar din păcate lucrările și codul publicat nu oferă aceste cele mai bune setări decât rareori.

GC-uri binecunoscute (de exemplu, DBpedia sau Wikidata) sunt updatate frecvent (de exemplu, lunar sau săptămânal). Acest lucru sugerează că o comparare a versiunii de astăzi cu un standard de aur creat acum câțiva ani poate să nu fie suficientă. Acest motiv a condus la pledoaria noastră pentru

importanța de a păstra evidențe despre versiunile GC utilizate pentru adnotarea seturilor de date, precum și a celor utilizate de adnotatori [WBKN19].

4.2 ROLUL INTERPRETĂRII ȘI EXPLICĂRII ÎN IA

4.2.1 Interpretare și explicare pentru librării agnostice de model

Spre deosebire de secțiunea despre benchmarking explicabil care s-a concentrat pe un un singur caz de utilizare, aici ne vom baza pe un sondaj publicat în [BA20b] și vom încerca să înțelegem tendințele generale din domeniu. Acest lucru ne va ajuta să mergem spre concluziile generale ale lucrării.

Programatorii cu cunoștințe statistice vor apăra termenul de interpretare în locul termenului de explicație. Pe de altă parte, termenul explicație poate fi preferat de artiști și designeri de vizualizare [BA20b]. Există desigur, unii cercetători din mijloc care pot folosi termenii în mod interschimbabil. Pentru cercetătorii NLP, putem susține că explicația ar trebui folosită și că dorim ca modelele noastre să ne ofere explicații clare în limbaj natural.

Oferirea unui răspuns clar cu privire la motivul pentru care un model a oferit o anumită predicție este o sarcină dificilă. În ceea ce privește ingineria, putem presupune că rezultatul este suma caracteristicilor modelului și să procedăm la deconstruirea contribuției fiecărei caracteristici. Aceasta a fost metoda implicită de câteva decenii [GE03].

În ultimul deceniu, lucrurile au evoluat. Bibliotecile moderne sunt construite în jurul ideei de model-agnosticism. Există o proclamație că astfel de librării software pot interpreta predicția oricărui model. Într-o anumită măsură acest lucru este posibil. De exemplu, pentru calculul valorilor Shapley, caracteristicile multiple sunt agregate și se calculează un scor care le reprezintă, urmând ca apoi să se calculeze contribuția din cadrul setului. Astfel de tehnici stau la baza bibliotecilor precum LIME [RSG16], SHAP [LL17] sau ELI5 [FJP + 19].

Principala critică împotriva acestor biblioteci se bazează pe trei idei: i) sunt ușor de utilizat, dar dificil de interpretat; ii) pot ajunge să culeagă efecte statistice în loc să compună interpretări sau explicații și (iii) atacurile adversariale împotriva lor pot fi propagate relativ ușor.

Ultima parte a criticii poate fi valabilă pentru orice fel de metodă de interpretare. Slack [SHJ + 20] demonstrează cum să comiți atacuri simple prin implementarea clasificatorilor corupți. Deoarece majoritatea modelelor sunt de fapt ansambluri, modificarea rezultatului unui singur clasificator poate duce la reducerea întregii construcții. Un singur clasificator corupt ar fi suficient pentru a adăuga bias față de o anumită categorie de oameni sau pentru a schimba un scor de credit. Atacurile robuste includ: crearea de falsuri în secvențele de jetoane care sunt concatenate cu șiruri și se pot transforma în

declanșatoare universale [WFK + 19]; îndepărtarea parțială a probelor de antrenament [CTW + 20]; ortografie aleatorie [SHY + 20]; și multe altele.

Statisticile oferă o alternativă la XAI: Neural Additive Models (NAM) care ar putea combina modele aditive generalizate (GAM) [AFZ + 20] cu caracteristici Deep Learning.

4.2.2 Explicarea rețelelor neurale recurente

Pentru a oferi o explicație bună a rezultatelor rețelelor neuronale este nevoie de ceva mai mult efort decât crearea unei vizualizări a stărilor ascunse. Cea mai bună idee este să urmărim fluxul de informații de la intrări la ieșiri, prin urmare să visualizăm setul de date de intrare, embeddings, capetele de atenție și diferite straturi ascunse, antrenamentul și rezultatul, de exemplu. Acest lucru necesită mult efort și foarte puține echipe au urmat această idee. În schimb, cazul mai frecvent implică vizualizarea unui singur subiect, cum ar fi embeddings (încorporări) sau capete de atenție.

Am descoperit trei grupuri mari de articole. Fiecare cluster conținea semnificativ mai multe lucrări decât cele discutate aici (de cel puțin zece ori mai multe lucrări, pe baza estimărilor noastre). Doar lucrări care au introdus concepte noi sau au avut numeroase citate au fost luate în considerare pentru includere.

Predicția următoarelor subiecte este un subiect clasic în NLP. Rezultatele sunt prezentate prin diagrame clasice (de exemplu, diagrame simple sau paralele, vizualizări matriciale).

Reprezentarea stărilor ascunse este subiectul central în explicarea RNN-urilor. Vizualizările incluse aici sunt aproape ca niște mini tablouri de bord. Toate includ un panou de control pentru navigarea între propoziții sau documente; un set de clustere de cuvinte sau activări neuronale; și vederi matriciale care sunt ideale pentru evidențierea rezultatelor. Mai multe sisteme care urmează aceste tipare includ: ActiVis [KAKC18], RNNVis [MCZ + 17] sau LSTMVis [SGPR18]. Proiectând astfel de interfețe este o activitate laborioasă de colaborare, prin urmare majoritatea lucrărilor includ o lungă listă de autori.

Rețelele de convoluție a graficelor (GCN) sunt incluse și ele aici. Există un număr mare de librării (de exemplu, PyTorch Geometric [FL19]), dar sunt axate pe câteva modele populare. Lucrările pe care le menționăm au fost publicate în ultimul an, dar ele arată că graficele și diagramele liniare sunt suficiente pentru aceste tipuri de modele.

4.2.3 Explicarea rețelelor Transformer

Transformatoarele au nevoie doar de atenție și de o pereche de codificatoare și decodificatoare și oferă rezultate mai bune decât media pentru multe sarcini. Este firesc să întrebăm cum este posibil

acest lucru? Este, de asemenea, firesc ca majoritatea vizualizărilor Transformer să se concentreze pe atenție. Ne concentrăm în schimb pe căutarea celor vizualizări care încearcă să ne ofere o înțelegere clară a întregul model, de la corpus, la hărți de atenție, la straturi neuronale și ieșiri multilingve.

Întrebarea centrală a lucrării care a lansat arhitectura Transformer: dacă atenția este în sine suficientă pentru rezolvarea multor sarcini [VSP + 17] este încă destul de controversată, în ciuda numărului mare de citări. Parțial această controversă este alimentată de numărul mare de GPU sau TPU necesare pentru faza de antrenament. În cele din urmă, dacă noi aruncăm toate resursele mai multor țări mici pentru a rezolva o singură problemă, sunt mari șanse să reușim, dar cu ce costuri pentru mediu sau pentru restul lumii [SGM19]? Ce se va întâmpla, de exemplu, cu țările sau companiile care au mai puține resurse? Cum vor concura? O metodă pentru a testa dacă atenția este suficientă, implică studierea efectelor unor greutăți manipulate pe ieșiri [JW19]. Dacă ieșirile sunt modificate, explicațiile captează probabil informațiile corecte. Acest lucru este echivalent cu studierea entropiei care trece prin sistem. O lucrare diferită sugerează că o astfel de tehnică ar trebui să fie aplicată numai în cazuri limită, de exemplu, în cazul în care instruirea contradictorie nu conduce la modificări grave în distribuția greutății [WP19]. Faimoasa lucrare despre papagalul stohastic [BGMS21] se învârte în jurul faptului că prejudecățile nu pot fi ușor eliminate din LM din cauza costurilor ridicate asociate cu recalificarea lor; dar această critică este infirmată de mulți cercetători care o consideră activism politic [Lis21]. În opinia noastră, prejudecățile pot fi identificate și eliminate în timp, dar procedeul poate fi costisitor.

Vizualizarea atenției oferă câteva informații, deci poate fi considerată o formă de explicație, chiar dacă nu întotdeauna granulară. Vizualizarea rețelelor Transformer este dedicată și încorporărilor și analizei dependenței [RYW + 19], dar și greutăților atenției în timpul pregătirii sau antrenamentului (de exemplu, [Vig19] sau [SZC + 20]), codificării fenomenelor lingvistice cum ar fi prepoziții sau corespondențe [CKLM19] sau sondelor structurale [HM19]. Așa cum a fost cazul RNN-urilor, vizualizarea stărilor ascunse ocupă, de asemenea, o cantitate semnificativă de literatură.

Putem discuta despre două categorii de vizualizări Transformer: (i) focalizate (sau subiect unic) și (ii) holistice (de exemplu, dedicată întregului flux de lucru sau model).

Unele subiecte de tendință în vizualizarea focalizată includ atenția (de exemplu, [AZ20], [VTM + 19], sau [Vig19]), sondarea [VST19], efectele interacțiunii informaționale ([HDWX20] și [VST19]) sau modele multilingve [TDP19].

Sondarea este considerată un tip special de explicație care prezintă informații lingvistice codificate în vectori [EER16]. Sondele structurale [HM19] reprezintă o versiune limitată a acestei probleme: testarea dacă arborele de sintaxă este încorporat în spațiul de reprezentare a cuvintelor unei rețele neuronale. Dacă se găsesc astfel de dovezi, atunci se poate presupune că geometria vectorială a

LM încorporează respectivii arbori de sintaxă. Criticii susțin că metoda funcționează bine pentru cazurile în care se cunosc distanțele cuvintelor, dar nu atunci când apar diferențe uriașe între acuratețea diverselor modele. De asemenea, proveniența LM-urilor contează la fel de mult Cu toate acestea, modelele bazate pe BERT pot dezvolta reprezentări lingvistice noi indiferent de originea lor comună. Voita [VT20] a sugerat că sondele ar trebui să transmită unele date (de exemplu, o descriere sau o etichetă) care pot fi evaluate pe baza lungimii sale. Mecanismul este stabil atunci când este implementat deasupra sondelor structurale.

Foarte puține vizualizări holistice includ datele de intrare (de exemplu, [SES + 20], și [HSG19]) sau dicționarele asociate [YCOL21], deși aceste erori se pot răspândi la sarcinile din aval [BRK + 18]. Erorile Transformer-ilor sunt examinate în [CKLM19]. Doar un mic subset de sisteme include vizualizări pentru toate componentele importante, inclusiv corpusul, încorporările, atenția și straturile ascunse (ExBERT [HSG19] și AttViz [SES + 20]).

Niciunul dintre sistemele examinate nu reușește să surprindă întreaga complexitate a unui sistem Transformer. Unul dintre principalele motive este concentrarea excesivă pe rolul atenției. Lipsa detaliilor despre codificatoare și decodificatoare este un alt motiv. Aceasta este o problemă complexă de proiectare. Combinarea atât a formei (de exemplu, arhitectura cu codificatoarele și decodificatoarele sale), cât și a funcției (de exemplu, căile informației neurale) într-o singură interfață este dificilă. Este tipic să te concentrezi numai pe un subiect. Forma este accentuată atunci când design-ul este axat pe circuite și logică, în timp ce funcția este accentuată atunci când proiectarea este focalizată asupra procesului. Pentru NLP, accentul pe funcție este suficient. Pentru a înțelege de ce aceste rețele funcționează atât de bine pentru diferite clase de probleme, inclusiv viziune, cel mai bine este să găsim un compromis între ambele sau, alternativ, să creăm două vizualizări separate.

4.2.4 Limbaj și viziune

Acesta este primul pas spre fuzionarea diferitelor ramuri ale IA, cum ar fi viziunea, NLP, vorbirea, SW sau robotica. În ultimii doi ani, acest subiect a devenit inevitabil la conferințele ML. Mai multe detalii despre aceste modele pot fi găsite în sondajul despre Visual Question Answering (VQA) [GCL + 20].

4.2.5 Discuție

Ultimele pagini au arătat faptul că, deși nu este clar dacă atenția este suficientă pentru a explica raționamentul sistemelor NLP, vizualizând atenția poate fi o cale spre explicații clare.

Design-ul modern al vizualizărilor DL [SGPR18] a fost stabilit în jurul momentului în care a fost publicată arhitectura Transformer [VSP + 17]. Ideea de bază a fost să împartă arhitectura pe baza

funcției și să se concentreze pe mai multe componente precum intrările, stările și ieșirile ascunse. Acest lucru a oferit o vizualizarea fluxului de lucru al informațiilor. Datorită publicației respective și adopției masive de care se bucură Transformers, au fost publicate mai multe vizualizări Transformers decât pentru celelalte arhitecturi. Aproape că se poate susține că Transformers a furnizat o soluție pentru majoritatea problemelor din NLP, dacă nu ar fi existat controversele legate de costul instruirii și de bias.

Ceea ce este clar este că majoritatea vizualizărilor sunt acum orientate spre model, în timp ce un cadru de vizualizare universal (sau cel puțin model-agnostic) pentru modelele neuronale nu există încă. O astfel de realizare ar deschide ușa către explicații universale. Nu este sigur că îi va convinge pe sceptici, dar merită construit. Nu dorim o înțelegere aproximativă a IA. Trebuie să înțelegem clar cum funcționează IA.

CAPITOLUL 5

CONCLUZIE ȘI MUNCĂ VIITOARE

5.1 IMPACT

Aceste capitole se bazează pe un set de publicații pentru conferințe și jurnale. Factorul de impact (IF) prezentat în tabele este în general pentru 2019 (publicat în 2020). Pentru conferințe, am luat în considerare clasamentul CORE din 2021 sau din ultimul an disponibil pentru conferința respectivă.

Capitolul 1 este introductiv.

Capitolul 2 a descris principalele avantaje și a prezentat unele aplicații ale grafurilor de cunoștințe. Un GC despre turism a fost publicat într-un articol de conferință la ENTER 2015 [SBÖ15] și într-un jurnal Springer *Journal of Information Technology and Tourism* [SOBS16]; tabloul de bord construit în jurul său fiind publicat in *Semantic Web Journal* [BSS + 17]. Detalii despre aceste publicații sunt incluse în Tabelul 5.1.

Tabelul 5.1 Impact pentru Capitolul 2

Articol	Outlet	Tip	Rang/IF
[SBÖ15]	ENTER 2015	Proceedings	C(2021)
[SOBS16]	Journal of IT & Tourism	Jurnal	IF=2.95
[BSS+17]	Semantic Web	Jurnal	IF=3.524

Capitolul 3 trece în revistă diverse contribuții legate de dezvoltarea sistemelor SAI. Primul set de contribuții dezvoltă ideea variantei numelui în sistemele de dizambiguizare a entităților, mai întâi prin algoritmi [WKB19], apoi prin lentile [BWN20] care pot ajuta la evaluarea rezultatelor. Aceste contribuții au fost construite în jurul unui sistem numit Recognyze, care este prezentat în [WKB18], în timp ce se menționează pe scurt o contribuție legată de un instrument de feliere a seturilor de date semantice [MSS + 17]. Al doilea set de contribuții dezvoltă ideea extinderii modelelor afective și construirea unor linii de bază rapide deasupra lor. O publicație despre această metodă a fost acceptată la *Cognitive Computation*, un prestigios jurnal Springer [WSB + 21]. Ultimul set de contribuții este dedicat verificării faptelor și este dedicat construirii unor linii de bază rapide folosind LM-uri bine cunoscute. Această contribuție a dus, de asemenea, la publicații de conferințe IWANN 2019 [BA19] și un articol de jurnal Springer în *Neural Proessing Letters* [BA20a].

Publicațiile rezumate în capitolul 3 sunt incluse în tabelul 5.2.

Tabelul 5.2 Impact pentru Capitolul 3

Articol	Outlet	Tip	Rang/IF
a) NEL			
[BNWS16]	LREC 2016	Proceedings	C(2021)
[MSS+17]	IEEE ICSC 2017	Proceedings	N/A
[WKB18]	ACM WIMS 2018	Conference	N/A
[BNW18]	ACM WIMS 2018	Proceedings	N/A
[WKB19]	LDK 2019	Proceedings	NEW(2019)
[BWN20]	ACL CoNLL 2020	Proceedings	A(2021)
b) Sentiment			
[WSB+21]	Cognitive Computation	Jurnal	IF=4.307
c) Verificare Fapte			
[BA19]	IWANN 2019	Proceedings	B(2018)
[BA20a]	Neural Processing Letters	Jurnal	IF=2.891

Capitolul 4 este construit în jurul ideii de explicabilitate. Primele trei secțiuni rezumă contribuțiile la evaluarea comparativă a sistemelor de dizambiguizae a entităților, inclusiv o taxonomie pentru analiza erorilor pulbicată la LREC 2018 [BRK + 18], un sistem de comparare evaluare vizuală numit Orbis și publicat la SEMANTICS 2018 [OKBW18], precum și câteva idei despre cum să îmbunătățim procesul de benchmarking pubicate la RANLP 2019 [WBKN19]. O contribuție legată de un subiect adiacent (extragerea conținutului forumurilor [WBWO21]) este, de asemenea menționată. Ultima parte a capitolului este dedicată unei anchete a explicabilității modelelor de limbaj și ajută la contextualizarea secțiunilor anterioare, sugerând totodată noi direcții de cercetare. Această contribuție a fost publicată în conferința IEEE IV 2020 [BA20b].

Contribuțiile discutate în capitolul 4 sunt incluse în tabelul 5.3.

Tabelul 5.3 Impact pentru Capitolul 4.

Articol	Outlet	Tip	Rang/IF
a) Benchmarking			
[BRK+18]	LREC 2018	Proceedings	C(2021)
[OKBW18]	SEMANTICS 2018	Proceedings	N/A(2021)
[WKB19]	ACL RANLP 2019	Proceedings	C(2021)
[WBWO21]	IEEE/WIC/ACM WI 2020	Proceedings	B(2021)
b) Visualizare			
[BA20b]	IEEE IV2020	Proceedings	B(2021)

Ultimul capitol trece în revistă contribuțiile și prin urmare nu conține citări la alte articole ale autorului.

5.2 CONCLUZIE

Această secțiune contextualizează și extinde secțiunile de discuții care au urmat fiecărei subcapitol al acestei lucrări.

Observațiile timpurii din capitolul 2 despre limitările clasice ale GC au servit drept platformă de lansare pentru restul contribuțiilor. Este clar că GC sunt utile și că trebuie să le completăm pentru a crea reprezentări bune.

Capitolul 3 a discutat trei aplicații NLP aparent fără legătură: dizambiguizarea entităților, verificarea sentimentului și a faptelor. În realitate, aceste aplicații sunt dependente una de alta și de aceea secvențele capitolului au fost aranjate în această ordine. Peste tot sunt necesare entități. Ele pot fi, de asemenea, utilizate în timpul calculului sentimentului și al verificării faptelor. În mod similar, exemplul special de verificare a faptelor discutat aici (detecția automată a știrilor false), are nevoie de entități și sentiment.

Este clar că problema variației numelui (capitolul 3.1) ar trebui să fie importantă pentru proiectarea sistemelor de dizambiguizare a entităților, cât și pentru sistemele de evaluare automată a acestei probleme. Ambele contribuții duc la mici îmbunătățiri în tratamentul variației numelui pentru sistemele de dezambiguizare (până la 2% pentru implementări algoritmice; până la 4-10% pentru lentile). Problema variației numelui este tratată în general prin intermediul potrivirilor parțiale în sisteme precum neleval [HNR14]. O singură lentilă poate fi considerată a fi echivalentă cu o potrivire parțială. O serie de lentile (de exemplu, ca cele prezentate în secțiunea 3.1.4), pe de altă parte, pot acoperi pe deplin toate cazurile de variație a numelui. Acesta este principalul motiv pentru care lentilele au fost dezvoltate - pentru a acoperi cât mai multe cazuri de variație.

Capitolul 3.2 acoperă o metodă de extindere a modelelor afective specifice unui domeniu în condiții de lipsă a resurselor. Evaluările ample au arătat că metoda funcționează. Scopul a fost să folosim metoda atât în mediul de cercetare, cât și în cel de producție. Teza a acoperit cazurile de utilizare a cercetării. Codul a fost ulterior adaptat și inclus în mediile de producție. Ceea ce este important de reținut este că modelele de limbaj au fost folosite ca parte a unui ansamblu mai mare care a inclus și GC, algoritmi de dizambiguizare a sensului și lexicoane de sentiment.

Capitolul 3.3 prezintă modul în care sistemele clasice sau Deep Learning pot fi utilizate pentru a detecta știri false. Întreaga secțiune arată că prin adăugarea mai multor atribute semantice cum ar fi entitățile sau sentimentul, este posibil să se obțină rezultate bune. Metoda este eficientă pentru crearea unor linii de bază rapide.

Metodele discutate în această secțiune au mai multe atribute în comun: i) toate folosesc GC și ML; ii) tratează probleme legate de un fel de variantă (de exemplu, varianța numelui la entități, adaptarea domeniului pentru sentiment, gradul de veridicitate pentru verificarea faptelor); și iii) toate conduc la linii de bază bune. Pentru a îmbunătăți aceste rezultate, pot fi imaginate arhitecturi mai sofisticate.

Capitolul 4 este axat pe explicabilitate. Sunt discutate contribuții multiple.

Capitolul 4.1 prezintă trei contribuții legate de benchmarking explicabil: i) o taxonomie a erorilor; ii) un instrument creat pentru vizualizarea benchmarking-ului; și iii) propunerea de a îmbunătăți publicarea seturilor de date folosite la evaluări prin includerea unor atribute suplimentare în metadatele lor. Toate aceste contribuții duc la o idee clară: mai multe lucruri trebuie să se facă pentru a îmbunătăți benchmarking-ul dezambiguizării entităților. Primii pași în această direcție au fost făcuți. Cu toate acestea, acest lucru poate fi realizat numai dacă întreaga comunitate este de acord să participe.

Capitolul 4.2 oferă o prezentare generală a stării actuale a vizualizărilor care ajută la explicarea modelelor de limbaj. Rezultatele sunt surprinzătoare. Multe progrese au fost realizate în ultimii 3-4 ani. Cu toate acestea, majoritatea vizualizărilor sunt focalizate pe funcționalitatea LM-urilor. Realizarea unui fel de echilibru între vizualizarea arhitecturilor și funcția acestora ar trebui să fie un țel important. Arhitectura pune în esență unele restricții asupra a ceea ce poate fi implementat. Informații suplimentare despre arhitectură (de exemplu, Ce operațiuni sunt acceptate? Cum sunt aceste operații vizualizate?) pot duce la o perspectivă interesantă. Această idee nu este explorată suficient încă.

Critica generală față de ML și, prin extensie, către IAS, se concentrează asupra problemelor dependențelor (de exemplu, de date sau dependență de domeniu), consistenței (de exemplu, transfer de cunoștințe, reglare fină) și transparenței (de exemplu, reproductibilitate) [CPGT17]. Această teză a oferit câteva idei despre cum să abordăm unele dintre aceste probleme. A descris cum să abordăm problema adaptabilității domeniului prin folosirea modelelor de limbaj împreună cu grafuri de cunoștințe. A abordat problema modificării modelelor existente în mod repetat prin discuțiile construite în jurul evaluărilor efectuate. Subiectul transparenței a fost discutat în contextul benchmarking-ului, precum și în contextul explicabilității. Acestea sunt doar câteva soluții posibile. Au funcționat pentru cazurile de utilizare studiate. Este posibil să nu funcționeze pentru alte cazuri de utilizare.

5.3 MUNCĂ VIITOARE

Există un interes reînnoit în proiectarea metodelor de vizualizare pentru explicarea rezultatelor rețelelor neuronale. Principala provocare va fi capturarea atât a arhitecturii cat și a funcțiilor rețelelor vizualizate.

O altă zonă interesantă de cercetare ar putea fi aplicarea NLP pentru seriile de timp. Aceasta poate include dezvoltarea de noi indicatori de sentiment și vizualizarea acestora.

Poate că cea mai importantă direcție viitoare de cercetare este să înțelegem cum pot supraviețui sistemele SAI fără o reprezentare externă a lumii (de exemplu, GC, hărți). După părerea noastră, sistemele AI semantice vor avea întotdeauna nevoie de interfețe prin care să-și salveze reprezentările.

BIBLIOGRAFIE

[ABC+16] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek Gordon Murray, Benoit Steiner, Paul A. Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zhang. Tensorflow: A System for LargeScale Machine Learning. CoRR, abs/1605.08695, 2016.

[AFZ+20] Rishabh Agarwal, Nicholas Frosst, Xuezhou Zhang, Rich Caruana, and Geoffrey E. Hinton. Neural additive models: Interpretable machine learning with neural nets. CoRR, abs/2004.13912, 2020.

[AG17] Hunt Allcott and Matthew Gentzkow. Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31(2):211–36, 2017.

[AOOV20] Tareq Al-Moslmi, Marc Gallofré Ocaña, Andreas L. Opdahl, andCsaba Veres. Named entity extraction for knowledge graphs: A literature overview. IEEE Access, 8:32862–32881, 2020.

[AS19] Heike Adel and Hinrich Schütze. Type-aware convolutional neural networks for slot filling. Journal of Artificial Intelligence Research, 66:297–339, 2019.

[AZ20] Samira Abnar and Willem H. Zuidema. Quantifying attention flow in transformers. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, pages 4190–4197. Association for Computational Linguistics, 2020.

[BA19] Adrian M.P. Brasoveanu and Razvan Andonie. Semantic fake news detection: A machine learning perspective. In Ignacio Rojas, Gonzalo Joya, and Andreu Català, editors, Advances in Computational Intelligence - 15th International Work-Conference on Artificial Neural Networks, IWANN 2019, Gran Canaria, Spain, June 12-14, 2019, Proceedings, Part I, volume 11506 of Lecture Notes in Computer Science, pages 656–667. Springer, 2019.

[BA20a] Adrian M.P. Brasoveanu and Razvan Andonie. Integrating machine learning techniques in semantic fake news detection. Neural Processing Letters, pages 1–18, 2020.

[BA20b] Adrian M.P. Brasoveanu and Razvan Andonie. Visualizing Transformers for NLP: A brief survey. In Ebad Banissi, Farzad Khosrowshahi, Anna Ursyn, Mark W. McK. Bannatyne, João Moura Pires, Nuno Datia, Kawa Nazemi, Boris Kovalerchuk, John Counsell, Andrew Agapiou, Zora Vrcelj, Hing-Wah Chau, Mengbi Li, Gehan Nagy, Richard Laing, Rita Francese, Muhammad Sarfraz, Fatma Bouali, Gilles

Venturini, Marjan Trutschl, Urska Cvek, Heimo Müller, Minoru Nakayama, Marco Temperini, Tania Di Mascio, Filippo Sciarrone, Veronica Rossano, Ralf Dörner, Loredana Caruccio, Autilia Vitiello, Weidong Huang, Michele Risi, Ugo Erra, Razvan Andonie, Muhammad Aurangzeb Ahmad, Ana Figueiras, and Mabule Samuel Mabakane, editors, 24th International Conference on Information Visualisation, IV 2020, Melbourne, Australia, September 7-11, 2020, pages 270–279. IEEE, 2020.

[BDS19] Jill Burstein, Christy Doran, and Thamar Solorio, editors. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). Association for Computational Linguistics, 2019.

[BGLL08] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10):P10008, 2008.

[BGMS21] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In Madeleine Clare Elish, William Isaac, and Richard S. Zemel, editors, FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021, pages 610–623. ACM, 2021.

[BNW18] Adrian M.P. Brasoveanu, Lyndon J.B. Nixon, and Albert Weichselbraun. Storylens: A multiple views corpus for location and event detection. In Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics (WIMS 2018), Novi Sad, Serbia, 2018. ACM.

[BNWS16] Adrian M.P. Brasoveanu, Lyndon J. B. Nixon, Albert Weichselbraun, and Arno Scharl. A regional news corpora for contextualized entity discovery and linking. In Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016., pages 3333–3338, 2016.

[BP06] Razvan C. Bunescu and Marius Pasca. Using encyclopedic knowledge for named entity disambiguation. In Diana McCarthy and Shuly Wintner, editors, EACL 2006, 11st Conference of theEuropean Chapter of the Association for Computational Linguistics, Proceedings of the Conference, April 3-7, 2006, Trento, Italy, pages 9–16. The Association for Computer Linguistics, 2006.

[BRK+18] Adrian M.P. Brasoveanu, Giuseppe Rizzo, Philipp Kuntschick, Albert Weichselbraun, and Lyndon J.B. Nixon. Framing named entity linking error types. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors,

Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), pages 266–271, Paris, France, may 2018. European Language Resources Association (ELRA).

[BSS+17] Adrian M.P. Brasoveanu, Marta Sabou, Arno Scharl, Alexander Hubmann-Haidvogel, and Daniel Fischl. Visualizing statistical linked knowledge for decision support. Semantic Web, 8(1):113–137, 2017.

[BWN20] Adrian M.P. Brasoveanu, Albert Weichselbraun, and Lyndon J. B. Nixon. In media res: A corpus for evaluating named entity linking with creative works. In Raquel Fernández and Tal Linzen, editors, Proceedings of the 24th Conference on Computational Natural Language Learning, CoNLL 2020, Online, November 19-20, 2020, pages 355–364. Association for Computational Linguistics, 2020.

[CCK+17] Diego Calvanese, Benjamin Cogrel, Sarah Komla-Ebri, Roman Kontchakov, Davide Lanti, Martin Rezk, Mariano Rodriguez-Muro, and Guohui Xiao. Ontop: Answering SPARQL queries over relational databases. Semantic Web, 8(3):471–487, 2017.

[CFC13] Marco Cornolti, Paolo Ferragina, and Massimiliano Ciaramita. A framework for benchmarking entity-annotation systems. In 22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, pages 249–260. International World Wide Web Conferences Steering Committee / ACM, 2013.

[Cho17] Francois Chollet. Deep Learning with Python. Manning Publications Co., 2017.

[CKLM19] Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. What does BERT look at? an analysis of bert's attention. CoRR, abs/1906.04341, 2019.

[CLH11] Erik Cambria, Andrew G. Livingstone, and Amir Hussain. The hourglass of emotions. In Anna Esposito, Antonietta Maria Esposito, Alessandro Vinciarelli, Rüdiger Hoffmann, and Vincent C. Müller, editors, Cognitive Behavioural Systems - COST 2102 International Training School, Dresden, Germany, February 21-26, 2011, Revised Selected Papers, volume 7403 of Lecture Notes in Computer Science, pages 144–157. Springer, 2011.

[CLX+20] Erik Cambria, Yang Li, Frank Z. Xing, Soujanya Poria, and Kenneth Kwok. SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis. In CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020, pages 105–114, 2020.

[CN16] Jason P. C. Chiu and Eric Nichols. Named Entity Recognition with Bidirectional LSTM-CNNs. TACL, 4:357–370, 2016.

[CNJA19] Ankush Chatterjee, Kedhar Nath Narahari, Meghana Joshi, and Puneet Agrawal. SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text. In Proceedings of the 13th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2019, Minneapolis, MN, USA, June 6-7, 2019, pages 39–48, 2019.

[CPGT17] Erik Cambria, Soujanya Poria, Alexander F. Gelbukh, and Mike Thelwall. Sentiment analysis is a big suitcase. IEEE Intell. Syst., 32(6):74–80, 2017.

[CTW+20] Nicholas Carlini, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom B. Brown, Dawn Song, Úlfar Erlingsson, Alina Oprea, and Colin Raffel. Extracting training data from large language models. CoRR, abs/2012.07805, 2020.

[DJHM13] Joachim Daiber, Max Jakob, Chris Hokamp, and Pablo N. Mendes. Improving efficiency and accuracy in multilingual entity extraction. In I-SEMANTICS 2013 - 9th International Conference on Semantic Systems, ISEM '13, Graz, Austria, September 4-6, 2013, pages 121–124. ACM, 2013.

[EER16] Allyson Ettinger, Ahmed Elgohary, and Philip Resnik. Probing for semantic evidence of composition by means of simple classification tasks. In Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP, RepEval@ACL 2016. Berlin, Germany, August 2016, pages 134–139. Association for Computational Linguistics, 2016.

[FJP+19] Angela Fan, Yacine Jernite, Ethan Perez, David Grangier, Jason Weston, and Michael Auli. ELI5: long form question answering. In Korhonen et al. [KTM19], pages 3558–3567.

[FK19] Haftu Wedajo Fentaw and Tae-Hyong Kim. Design and investigation of capsule networks for sentence classification. Applied Sciences, 9(11):2200, 2019.

[FL19] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. CoRR, abs/1903.02428, 2019.

[GCL+20] Zhe Gan, Yen-Chun Chen, Linjie Li, Chen Zhu, Yu Cheng, and Jingjing Liu. Large-scale adversarial training for visionand-language representation learning. In Hugo Larochelle,

Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020.

[GE03] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. J. Mach. Learn. Res., 3:1157–1182, 2003.

[GvLB+17] Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors. Advances in Neural Information Processing Systems 30:

Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA, 2017.

[HBC+20] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutiérrez, José Emilio Labra Gayo, Sabrina Kirrane, Sebastian Neumaier, Axel Polleres, Roberto Navigli, Axel-Cyrille Ngonga Ngomo, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan F. Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge graphs. CoRR, abs/2003.02320, 2020.

[HDWX20] Yaru Hao, Li Dong, Furu Wei, and Ke Xu. Self-attention attribution: Interpreting information interactions inside transformer. CoRR, abs/2004.11207, 2020.

[HLAN12] Sebastian Hellmann, Jens Lehmann, Sören Auer, and Marcus Nitzschke. NIF combinator: Combining NLP tool output. In Knowledge Engineering and Knowledge Management - 18th International Conference, EKAW 2012, Galway City, Ireland, October 8-12, 2012. Proceedings, volume 7603 of Lecture Notes in Computer Science, pages 446–449. Springer, 2012.

[HM19] John Hewitt and Christopher D. Manning. A structural probe for finding syntax in word representations. In Burstein et al. [BDS19], pages 4129–4138.

[HNR14] Ben Hachey, Joel Nothman, and Will Radford. Cheap and easy entity evaluation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22- 27, 2014, Baltimore, MD, USA, Volume 2: Short Papers, pages 464–469. The Association for Computer Linguistics, 2014.

[HSG19] Benjamin Hoover, Hendrik Strobelt, and Sebastian Gehrmann. exbert: A visual analysis tool to explore learned representations in transformers models. CoRR, abs/1910.05276, 2019.

[HSN+12] Johannes Hoffart, Stephan Seufert, Dat Ba Nguyen, Martin Theobald, and Gerhard Weikum. KORE: keyphrase overlap relatedness for entity disambiguation. In 21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012, pages 545–554. ACM, 2012.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition. Springer Series in Statistics. Springer, 2009.

[HYB+11] Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. Robust disambiguation of named entities in text. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language

Processing, EMNLP 2011, 27-31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 782–792, 2011.

[IJNW19] Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019. Association for Computational Linguistics, 2019.

[ITA+16] Kazuki Irie, Zoltán Tüske, Tamer Alkhouli, Ralf Schlüter, and Hermann Ney. LSTM, GRU, Highway and a Bit of Attention: An Empirical Overview for Language Modeling in Speech Recognition. In Nelson Morgan, editor, Interspeech 2016, 17th Annual Conference of the International Speech Communication Association, San Francisco, CA, USA, September 8-12, 2016, pages 3519–3523. ISCA, 2016.

[JdMAJ+20] Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T Evelo, et al. Fair principles: interpretations and implementation considerations, 2020.

[JRN17] Kunal Jha, Michael Röder, and Axel-Cyrille Ngonga Ngomo. All that glitters is not gold - rule-based curation of reference datasets for named entity recognition and entity linking. In Eva Blomqvist, Diana Maynard, Aldo Gangemi, Rinke Hoekstra, Pascal Hitzler, and Olaf Hartig, editors, The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I, volume 10249 of Lecture Notes in Computer Science, pages 305–320, 2017.

[JW19] Sarthak Jain and Byron C. Wallace. Attention is not explanation. In Burstein et al. [BDS19], pages 3543–3556.

[KAKC18] Minsuk Kahng, Pierre Y. Andrews, Aditya Kalro, and Duen Horng (Polo) Chau. Activis: Visual exploration of industry-scale deep neural network models. IEEE Trans. Vis. Comput. Graph., 24(1):88–97, 2018.

[KB14] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980, 2014.

[Kim14] Yoon Kim. Convolutional neural networks for sentence classification. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25- 29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 1746–1751. ACL, 2014.

[KJPC20] Jaeyoung Kim, Sion Jang, Eunjeong L. Park, and Sungchul Choi. Text classification using capsules. Neurocomputing, 376:214–221, 2020.

[KTM19] Anna Korhonen, David R. Traum, and Lluís Màrquez, editors. Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers. Association for Computational Linguistics, 2019.

[LEB12] Günther Lengauer, Frank Esser, and Rosa Berganza. Negativity in Political News: A Review of Concepts, Operationalizations and Key Findings. Journalism, 13(2):179–202, 2012.

[LIJ+15] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. Dbpedia - A large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web, 6(2):167–195, 2015.

[Lis21] Michael Lissack. The slodderwetenschap (sloppy science) of stochastic parrots - A plea for science to NOT take the route advocated by gebru and bender. CoRR, abs/2101.10098, 2021.

[LL17] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In Guyon et al. [GvLB+17], pages 4765–4774.

[LLX+17] Yunfei Long, Qin Lu, Rong Xiang, Minglei Li, and Chu-Ren Huang. Fake News Detection Through Multi-Perspective Speaker Profiles. In Greg Kondrak and Taro Watanabe, editors, Proceedings of the Eighth International Joint Conference on Natural Language Processing, IJCNLP 2017, Taipei, Taiwan, November 27 - December 1, 2017, Volume 2: Short Papers, pages 252–256. Asian Federation of Natural Language Processing, 2017.

[MCB+20] Giovanni Da San Martino, Stefano Cresci, Alberto BarrónCedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. A survey on computational propaganda detection. In Christian Bessiere, editor, Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, pages 4826–4832. ijcai.org, 2020. Scheduled for July 2020, Yokohama, Japan, postponed due to the Corona pandemic.

[MCZ+17] Yao Ming, Shaozu Cao, Ruixiang Zhang, Zhen Li, Yuanzhe Chen, Yangqiu Song, and Huamin Qu. Understanding hidden memories of recurrent neural networks. In Brian D. Fisher, Shixia Liu, and Tobias Schreck, editors, 12th IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2017, Phoenix,AZ, USA, October 3-6, 2017, pages 13–24. IEEE Computer Society, 2017.

[MRN14a] Andrea Moro, Alessandro Raganato, and Roberto Navigli. Entity linking meets word sense disambiguation: a unified approach. TACL, 2:231–244, 2014.

[MRN14b] Andrea Moro, Alessandro Raganato, and Roberto Navigli. Entity linking meets word sense disambiguation: a unified approach. Transactions of the Association for Computational Linguistics, 2:231–244, 2014.

[MSS+17] Edgard Marx, Saeedeh Shekarpour, Tommaso Soru, Adrian M.P. Brasoveanu, Muhammad Saleem, Ciro Baron, Albert Weichselbraun, Jens Lehmann, Axel-Cyrille Ngonga Ngomo, and Sören Auer. Torpedo: Improving the state-of-the-art RDF dataset slicing. In 11th IEEE International Conference on Semantic Computing, ICSC 2017, San Diego, CA, USA, January 30 - February 1, 2017, pages 149–156, San Diego, CA, USA, 2017. IEEE Computer Society.

[NGP+15] Andrea Giovanni Nuzzolese, Anna Lisa Gentile, Valentina Presutti, Aldo Gangemi, Darío Garigliotti, and Roberto Navigli. Open knowledge extraction challenge. In Semantic Web Evaluation Challenges - Second SemWebEval Challenge at ESWC 2015, Portorož, Slovenia, May 31 - June 4, 2015, Revised Selected Papers, volume 548 of Communications in Computer and Information Science, pages 3–15. Springer, 2015.

[OKBW18] Fabian Odoni, Philipp Kuntschik, Adrian M.P. Brasoveanu, and Albert Weichselbraun. On the importance of drill-down analysis for assessing gold standards and named entity linking performance. In Anna Fensel, Victor de Boer, Tassilo Pellegrini, Elmar Kiesling, Bernhard Haslhofer, Laura Hollink, and Alexander Schindler, editors, Proceedings of the 14th International Conference on Semantic Systems, SEMANTICS 2018, Vienna, Austria, September 10-13, 2018, volume 137 of Procedia Computer Science, pages 33–42. Elsevier, 2018.

[OMK21] Daniel W. Otter, Julian R. Medina, and Jugal K. Kalita. A survey of the usages of deep learning for natural language processing. IEEE Trans. Neural Networks Learn. Syst., 32(2):604–624, 2021.

[PCLG21] Deepak P, Tanmoy Chakraborty, Cheng Long, and Santhosh Kumar G. Data Science for Fake News - Surveys and Perspectives, volume 42 of The Information Retrieval Series. Springer, 2021.

[PHM+19] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, pages 527–536, 2019.

[PRT16] Julien Plu, Giuseppe Rizzo, and Raphaël Troncy. Enhancing entity linking by combining NER models. In Harald Sack, Stefan Dietze, Anna Tordai, and Christoph Lange, editors, Semantic Web Challenges - Third SemWebEval Challenge at ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers, volume 641 of Communications in Computer

and Information Science, pages 17–32. Springer, 2016.

[PZS+20] Beatrice Portelli, Jason Zhao, Tal Schuster, Giuseppe Serra, and Enrico Santus. Distilling the evidence to augment fact verification models. In Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER), pages 47–51, 2020.

[QSF+18] Ye Qi, Devendra Singh Sachan, Matthieu Felix, Sarguna Padmanabhan, and Graham Neubig. When and why are pre-trained word embeddings useful for neural machine translation? In Marilyn A. Walker, Heng Ji, and Amanda Stent, editors, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers), pages 529–535. Association for Computational Linguistics, 2018.

[Rad15] Will Radford. Linking Named Entities to Wikipedia. PhD thesis, School of Information Technologies, Faculty of Engineering and IT, The University of Sydney, 2015.

[RCJ+17] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. Truth of varying shades: Analyzing language in fake news and political fact-checking. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017, pages 2931–2937. Association for Computational Linguistics, 2017.

[RS20] Maithra Raghu and Eric Schmidt. A survey of deep learning for scientific discovery. CoRR, abs/2003.11755, 2020.

[RSG16] Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?": Explaining the predictions of any classifier. In Balaji Krishnapuram, Mohak Shah, Alexander J. Smola, Charu C. Aggarwal, Dou Shen, and Rajeev Rastogi, editors, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13- 17, 2016, pages 1135–1144. ACM, 2016.

[RUH+14] Michael Röder, Ricardo Usbeck, Sebastian Hellmann, Daniel Gerber, and Andreas Both. N3

- A collection of datasets for named entity recognition and disambiguation in the NLP interchange format. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014, pages 3529–3533. European Language Resources Association (ELRA), 2014.

[RUN18] Michael Röder, Ricardo Usbeck, and Axel-Cyrille Ngonga Ngomo. GERBIL - benchmarking named entity recognition and linking consistently. Semantic Web, 9(5):605–625, 2018.

[RYW+19] Emily Reif, Ann Yuan, Martin Wattenberg, Fernanda B. Viégas, Andy Coenen, Adam Pearce, and Been Kim. Visualizing and measuring the geometry of BERT. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada, pages 8592–8600, 2019.

[SBF+18] Abeed Sarker, Maksim Belousov, Jasper Friedrichs, Kai Hakala, Svetlana Kiritchenko, Farrokh Mehryary, Sifei Han, Tung Tran, Anthony Rios, Ramakanth Kavuluru, Berry de Bruijn, Filip Ginter, Debanjan Mahata, Saif M. Mohammad, Goran Nenadic, and Graciela Gonzalez-Hernandez. Data and systems for medicationrelated text classification and concept normalization from twitter: insights from the social media mining for health (SMM4H)-2017 shared task. Journal of American Medical Informatics Association, 25(10):1274–1283, 2018.

[SBÖ15] Marta Sabou, Adrian M.P. Brasoveanu, and Irem Önder. Linked data for cross-domain decision-making in tourism. In Iis Tussyadiah and Alessandro Inversini, editors, Information and Communication Technologies in Tourism 2015, ENTER 2015, Proceedings of the International Conference in Lugano, Switzerland, February 3 - 6, 2015, pages 197–210. Springer, 2015.

[SCH17] Robyn Speer, Joshua Chin, and Catherine Havasi. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA, pages 4444–4451, 2017.

[SES+20] Blaz Skrlj, Nika Erzen, Shane Sheehan, Saturnino Luz, Marko Robnik-Sikonja, and Senja Pollak. Attviz: Online exploration of self-attention for transparent neural language modeling. CoRR, abs/2005.05716, 2020.

[SGM19] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in NLP. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, pages 3645–3650. Association forComputational Linguistics, 2019.

[SGPR18] Hendrik Strobelt, Sebastian Gehrmann, Hanspeter Pfister, and Alexander M. Rush. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. IEEE Trans. Vis. Comput. Graph., 24(1):667–676, 2018.

[SHJ+20] Dylan Slack, Sophie Hilgard, Emily Jia, Sameer Singh, and Himabindu Lakkaraju. Fooling LIME and SHAP: adversarial attacks on post hoc explanation methods. In Annette N. Markham, Julia Powles,

Toby Walsh, and Anne L. Washington, editors, AIES '20: AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA, February 7-8, 2020, pages 180–186. ACM, 2020.

[SHY+20] Lichao Sun, Kazuma Hashimoto, Wenpeng Yin, Akari Asai, Jia Li, Philip S. Yu, and Caiming Xiong. Adv-bert: BERT is not robust on misspellings! generating nature adversarial samples on BERT. CoRR, abs/2003.04985, 2020.

[SLCC20] Yosephine Susanto, Andrew G. Livingstone, Ng Bee Chin, and Erik Cambria. The hourglass model revisited. IEEE Intell. Syst., 35(5):96–102, 2020.

[SLH+18] Elvis Saravia, Hsien-Chi Toby Liu, Yen-Hao Huang, Junlin Wu, and Yi-Shin Chen. CARER: Contextualized Affect Representations for Emotion Recognition. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018, pages 3687– 3697, 2018.

[SM03] Erik F. Tjong Kim Sang and Fien De Meulder. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In Walter Daelemans and Miles Osborne, editors, Proceedings of the Seventh Conference on Natural Language Learning, CoNLL 2003, Held in cooperation with HLT-NAACL 2003, Edmonton, Canada, May 31 - June 1, 2003, pages 142–147. ACL, 2003.

[SOBS16] Marta Sabou, Irem Onder, Adrian M.P. Brasoveanu, and Arno Scharl. Towards cross-domain data analytics in tourism: a linked data based approach. J. Inf. Technol. Tour., 16(1):71–101, 2016.

[SZC+20] Weijie Su, Xizhou Zhu, Yue Cao, Bin Li, Lewei Lu, Furu Wei, and Jifeng Dai. VL-BERT: pre-training of generic visual-linguistic representations. In 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net, 2020.

[TDP19] Ian Tenney, Dipanjan Das, and Ellie Pavlick. BERT rediscovers the classical NLP pipeline. In Korhonen et al. [KTM19], pages 4593–4601.

[TVCM18] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. FEVER: a large-scale dataset for fact extraction and verification. In Marilyn A. Walker, Heng Ji, and Amanda Stent, editors, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers), pages 809–819. Association for Computational Linguistics, 2018.

[vEMP+16] Marieke van Erp, Pablo N. Mendes, Heiko Paulheim, Filip Ilievski, Julien Plu, Giuseppe Rizzo, and Jörg Waitelonis. Evaluating entity linking: An analysis of current benchmark datasets and a

roadmap for doing a better job. In Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016., pages 4373– 4379, 2016.

[Vig19] Jesse Vig. A multiscale visualization of attention in the transformer model. In Marta R. Costa-jussà and Enrique Alfonseca, editors, Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28 - August 2, 2019, Volume 3: System Demonstrations, pages 37– 42. Association for Computational Linguistics, 2019.

[Vra13] Denny Vrandecic. The rise of wikidata. IEEE Intelligent Systems, 28(4):90–95, 2013.

[VSP+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Guyon et al. [GvLB+17], pages 5998–

6008.

[VST19] Elena Voita, Rico Sennrich, and Ivan Titov. The bottom-up evolution of representations in the transformer: A study with machine translation and language modeling objectives. In Inui et al. [IJNW19], pages 4395–4405.

[VT20] Elena Voita and Ivan Titov. Information-theoretic probing with minimum description length. CoRR, abs/2003.12298, 2020.

[VTM+19] Elena Voita, David Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. In Korhonen et al. [KTM19], pages 5797–5808.

[Wan17] William Yang Wang. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. CoRR, abs/1705.00648, 2017.

[WBKN19] Albert Weichselbraun, Adrian M.P. Braşoveanu, Philipp Kuntschik, and Lyndon J.B. Nixon. Improving named entity linking corpora quality. RANLP 2019, pages 1328–1337, 2019.

[WBWO21] Albert Weichselbraun, Adrian M.P. Brasoveanu, Roger Waldvogel, and Fabian Odoni. Harvest - an open source toolkit for extracting posts and post metadata from web forums. CoRR, abs/2102.02240, 2021.

[WFK+19] Eric Wallace, Shi Feng, Nikhil Kandpal, Matt Gardner, and Sameer Singh. Universal adversarial triggers for attacking and analyzing NLP. In Inui et al. [IJNW19], pages 2153–2162.

[WKB18] Albert Weichselbraun, Philipp Kuntschik, and Adrian M.P. Brasoveanu. Mining and leveraging background knowledge for improving named entity linking. In Proceedins of the 8th International

Conference on Web Intelligence, Mining and Semantics (WIMS 2018), pages 27:1–27:11, Novi Sad, Serbia, 2018. ACM.

[WKB19] Albert Weichselbraun, Philipp Kuntschik, and Adrian M.P. Brasoveanu. Name variants for improving entity discovery and linking. In Maria Eskevich, Gerard de Melo, Christian Fäth, John P. McCrae, Paul Buitelaar, Christian Chiarcos, Bettina Klimek, and Milan Dojchinovski, editors, 2nd Conference on Language, Data and Knowledge, LDK 2019, May 20-23, 2019, Leipzig, Germany., volume 70 of OASICS, pages 14:1–14:15. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2019.

[Wöb03] Karl W Wöber. Information supply in tourism management by marketing decision support systems. Tourism Management, 24(3):241–255, 2003.

[WP19] Sarah Wiegreffe and Yuval Pinter. Attention is not not explanation. In Inui et al. [IJNW19], pages 11–20.

[WSB+21] Albert Weichselbraun, Jakob Steixner, Adrian MP Braşoveanu, Arno Scharl, Max Göbel, and Lyndon JB Nixon. Automatic expansion of domain-specific affective models for web intelligence applications. Cognitive Computation, pages 1–18, 2021.

[YCOL21] Zeyu Yun, Yubei Chen, Bruno A. Olshausen, and Yann LeCun. Transformer visualization via dictionary learning: contextualized embedding as a linear superposition of transformer factors. CoRR, abs/2103.15949, 2021.

[YHPC18] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing [review article]. IEEE Comp. Int. Mag., 13(3):55–75, 2018.

Adrian M.P. Braşoveanu
Lista de publicaţii

Articole publicate în jurnale indexate ISI

1. Weichselbraun, A., Steixner, J., **Braşoveanu, A.M.P.**, Scharl, A., Gobel, M., Nixon, L.B. (2021). Automatic Expansion of Domain-Specific Affective Models for Web Intelligence Applications. Cognitive Computation 13. DOI: 10.1007/s12559-021-09839-4. (**IF=4.307**)

2. **Braşoveanu, A.M.P.**, Andonie, R. (2020). Integrating Machine Learning Techniques in Semantic Fake News Detection. Neural Processing Letters, 1-18. DOI: 10.1007/s11063-020-10365-x. (**IF=2.891**)

3. **Braşoveanu, A.M.P.**, M. Sabou, Scharl, A. Hubmann-Haidvogel, A., Fischl, D. (2017). Visualizing Statistical Linked Knowledge for Decision Support. Semantic Web, 8.1, pp. 113137. DOI: 10.3233/SW-160225. (**IF=3.524**)

4. Sabou, M., Onder, I., **Braşoveanu, A.M.P.**, Scharl, A. (2016). Towards Cross-Domain Data Analytics in Tourism: A Linked Data Based Approach. J. of IT & Tourism 16.1, pp. 71-101. DOI: 10.1007/s40558-015-0049-5. (**IF=2.95**)

Articole în conferinţe indexate în bazele de date internaţionale

5. **Braşoveanu, A.M.P.,** Weichselbraun, A., Nixon, L. (2020). In Media Res: A Corpus for Evaluating Named Entity Linking with Creative Works. In Proceedings of the 24th Conference on Computational Natural Language Learning, ACL, pp. 355-364. DOI: 10.18653/v1/2020.conll-1.28.

6. Weichselbraun, A., **Braşoveanu, A.M.P.**, Waldvogel, R., Odoni, F. (2021). Harvest - An Open Source Toolkit for Extracting Posts and Post Metadata from Web Forums. IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Melbourne, Australia, IEEE/ACM, pp. 430-436. DOI: https://arxiv.org/abs/2102.02240.

7. **Braşoveanu, A.M.P.**, Andonie, R. (2020) Visualizing Transformers for NLP: A Brief Survey. 24th International Conference on Information Visualisation, IV 2020, Melbourne, Australia, IEEE, pp. 270-279. DOI: 10.1109/IV51561.2020. 00051.

8. Weichselbraun, A., **Braşoveanu, A.M.P.**, Kuntschik, P., Nixon, L.J.B. (2019). Improving Named Entity Linking Corpora Quality. RANLP 2019, Varna, Bulgaria, Incoma, pp. 1329-1338. DOI: 10.26615/978-954-452-056-4_152.

9. Weichselbraun, A., Kuntschik, P., **Braşoveanu, A.M.P.** (2019). Name Variants for Improving Entity Discovery and Linking. LDK 2019, Leipzig, Germany. OASICS 70, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik 2019, pp. 14:1-14:15. DOI: 10.4230/OASIcs.LDK.2019.14.

10. **Braşoveanu, A.M.P.**, Andonie, R. (2019). Semantic Fake News Detection: A Machine Learning Perspective. IWANN 2019 Part I, Springer, pp. 656-667. DOI: 10.1007/978-3-030-20521-8_54.

11. Odoni, F., **Braşoveanu, A.M.P.**, Kuntschik, P., Weichselbraun, A. (2019). Introducing orbis: An extendable evaluation pipeline for named entity linking performance drilldown analyses. Proceedings of the Association for Information Science and Technology, 56(1), 468-471. DOI: 10.1002/pra2.49.

12. Odoni, F., Kuntschik, P., **Braşoveanu, A.M.P.**, Weichselbraun, A. (2018). On the Importance of Drill-Down Analysis for Assessing Gold Standards and Named Entity Linking Performance. Semantics 2018, In Procedia Computer Science, 137, 33-42. Elsevier. DOI: 10.1016/j.procs.2018.09.004.

13. **Braşoveanu, A.M.P.**, Rizzo, G., Kuntschik, P., Weichselbraun, A., Nixon, L. (2018). Framing Named Entity Linking Error Types. LREC 2018, Miyzaki, Japan, pp. 266-271. ELRA, Paris, France. DOI: www.lrecconf.org/proceedings/lrec2018/pdf/612.pdf.

14. Weichselbraun, A., Kuntschik, P., **Braşoveanu, A.M.P.** (2018). Mining and Leveraging Background Knowledge for Improving Named Entity Linking. WIMS 2018, Novi Sad, Serbia, ACM, pp. 27:1-27:11. DOI: 10.1145/3227609.3227670.

15. **Braşoveanu, A.M.P.**, Nixon, L. J., Weichselbraun, A. (2018). StoryLens: A multiple views corpus for location and event detection. WIMS 2018, Novi Sad, Serbia, ACM, pp. 30:1-30:4. DOI: 10.1145/3227609.3227674.

16. Marx, E., Sherkarpour, S., Soru, T., **Braşoveanu, A.M.P.**, Saleem, M., Baron, C., Weichselbraun, A., Lehmann, J., Ngonga Ngomo, A.-C., Auer, S. (2017). Torpedo: Improving the State-of-the-Art RDF Dataset Slicing. ICSC 2017, San Diego, California, IEEE, pp. 149-156. DOI: 10.1109/ICSC.2017.79.

17. **Braşoveanu, A.M.P.**, L.J.B. Nixon, A. Weichselbraun, A. Scharl. A (2016) Regional News Corpora for Contextualized Entity Discovery and Linking. LREC 2016, ELRA, pp. 3333-3338. DOI: www.lrecconf.org/proceedings/lrec2016/summaries/835.html.

18. Sabou, M., **Braşoveanu, A.M.P.**, Onder, I. (2015). Linked Data for Cross-Domain Decision-Making in Tourism. ENTER 2015, Springer, pp. 197-210. DOI: 10.1007/978-3-319-14343-9_15.

Short Abstract - Scurt Rezumat

Abstract. Semantic AI is a recent approach towards AI that is focused on combining semantics with classic AI methods like classification or clustering. By adding semantics, we can increase data quality while removing black-box approaches. Its core proposition is that regardless of its original provenance, data can be processed and stored into refined formats like those provided by knowledge graphs or search engines. These open data clusters can later be used to solve complex problems with hybrid approaches. By combining entities extracted from a KG with sentiment and ML classifiers, it is possible to verify the claims from a sentence, for example. This thesis examines several hybrid methods enabled by SAI to understand how to leverage them to build baselines for research and production. Once these methods are examined, it emerges that each component may add its errors to the stack and confuse the researchers and developers. It then argues that to move forward, it is important to build some practical solutions like a taxonomy of errors or a tool for visualizing benchmarking results, to help researchers navigate this complexity.

Rezumat. IA semantică este o abordare recentă de IA prin care se combină semantica şi metodele clasice de IA, cum ar fi clasificarea sau clusterizarea. Adăugând semantică, putem creşte calitatea datelor, eliminând în acelaşi timp abordările de tip cutie neagră. Propunerea ei de bază este că, indiferent de provenienţa originală, datele pot fi procesate şi stocate în formate rafinate, precum cele furnizate de reţele semantice sau motoare de căutare. Aceste clustere de date pot fi utilizate ulterior pentru a rezolva probleme complexe cu abordări hibride. Combinând entităţi extrase dintr-o reţea semantică cu analiza sentimentului şi clasificatori bazaţi pe învăţare automată, este posibil să se verifice afirmaţiile dintr-o propoziţie, de exemplu. Această teză examinează mai multe metode hibride propuse de IA semantică pentru a înţelege cum să le folosească pentru a construi linii de bază pentru cercetare şi producţie. Odată examinate aceste metode, rezultă că fiecare componentă îşi poate adăuga erorile în stivă şi poate deruta cercetătorii şi dezvoltatorii. Apoi argumentează că pentru a merge mai departe, este important să construim câteva soluţii practice, cum ar fi o taxonomie a erorilor sau un instrument pentru vizualizarea rezultatelor evaluărilor, pentru a ajuta cercetătorii să navigheze această complexitate.

Adrian M.P. Braşoveanu, M.Sc. - CV

Main Areas of Research	Natural Language Processing (NLP) Semantic Web (SW) and Knowledge Graphs (KG) Machine Learning (ML) Information Visualization (IV)

Education

2014-Present	**PhD Student**, Transylvania University, Braşov, România. Domain: Computer Science. Thesis topic: *Intelligent Systems in Semantic Networks*.
2008	**MSc in Distributed and Parallel Processing Systems (Dipl.-Ing.),** Lucian Blaga University of Sibiu, Sibiu, România
2007	**MSc in Computer Science and Automatic Control (Dipl.-Ing.),** Lucian Blaga University of Sibiu, Sibiu, România

Work Experience

2020.11-Present	**Researcher**, MODUL University Vienna, Austria
2018.11- Present	**Researcher**, MODUL Technology GmbH, Vienna, Austria
2017.11-2018.10	**Invited Researcher**, University of Applied Sciences of the Grisons (UASG), Chur, Switzerland
2016.06-2017.10	**Researcher**, MODUL Technology GmbH, Vienna, Austria.
2011.11- 2016.05	**Researcher,** MODUL University Vienna, Austria.
2011.03- 2011.10	**Intern**, MODUL University, Austria
2007.09- 2011.02	**Java Programmer**, em2Soft, Sibiu, Romania.

Key International Cooperation Partners

LINKS	LINKS Foundations, Torino (Italy), Senior Researcher Giuseppe Rizzo
AKSW	AKSW Leipzig (Germany), Senior Researcher Milan Dojchinovski
DICE	DICE Lab at Paderborn University (Germany), Prof. Axel-Cyrille Ngonga Ngomo

Most Important Research Projects

ReTV	European Horizon 2020 Programme (Researcher)
InVID	European Horizon 2020 Programme (Researcher)
ASAP	EU 7th Framework Program (Researcher)
LinkedTV	EU 7th Framework Program (Researcher)
ETIHQ	A PlanetData WP - EU 7th Framework Program (Researcher)

Awards

ISWC 2017	**Best Reviewer Award** at the **International Semantic Web Conference 2017 (ISWC 2017)**
IEEE ICSC 2017	**Honorable Mention Award**, Awarded for E. Marx, S. Sherkarpour, T. Soru, **A.M.P. Braşoveanu**, M. Saleem, C. Baron, A. Weichselbraun, J. Lehmann, A.-C. Ngonga Ngomo, S. Auer. Torpedo: Improving the State-of-the-Art RDF Dataset Slicing.

List of Publications

Journal Articles

- Weichselbraun, A., Steixner, J., **Braşoveanu, A.M.P.**, Scharl, A., Gobel, M., Nixon, L.B.J. (2021). Automatic Expansion of Domain-Specific Affective Models for Web Intelligence Applications. *Cognitive Computation* 13. 10.1007/s12559-021-09839-4.
- **Braşoveanu, A.M.P.**, Andonie, R. (2020). Integrating Machine Learning Techniques in Semantic Fake News Detection. Neural Processing Letters, 1-18. DOI: 10.1007/s11063-020-10365-x.
- **Braşoveanu, A.M.P.**, Sabou, M., Scharl, A., Hubmann-Haidvogel, A., Fischl, D. (2017). Visualizing Statistical Linked Knowledge for Decision Support. Semantic Web 8(1), pp. 113–137. DOI: 10.3233/SW-160225.
- Sabou, M., Onder, I., **Braşoveanu, A.M.P.**, Scharl, A (2016). Towards Cross-Domain Data Analytics in Tourism: A Linked Data Based Approach. J. of IT & Tourism 16(1), pp. 71-101. DOI: 10.1007/s40558-015-0049-5.
- Sabou, M., Arsal, I., **Braşoveanu, A.M.P.** (2013). TourMISLOD: A tourism linked data set. Semantic Web 4(3), pp. 271-276. DOI: 10.3233/SW-2012-0087.
- **Braşoveanu, A.M.P**, Dzitac, I (2012). The Role of Visual Rhetoric in Semantic Multimedia: Strategies for Decision Making in Times of Crisis. Int. J. Comput. Commun. Control, 7(4), pp. 606-616. DOI: 10.15837/ijccc.2012.4.1361.
- **Braşoveanu, A.M.P.**, Nagy, M, Mateut-Petrisor, O., Urziceanu, R. (2010). The Avatar in the Context of Intelligent Social Semantic Web. Int. J. Comput. Commun. Control, 5(4), pp. 477-482. DOI: 10.15837/ijccc.2010.4.2497.
- **Braşoveanu, A.M.P.**, Manolescu, A., Spinu, M.N. (2010). Generic Multimodal Ontologies for Human-Agent Interaction. Int. J. Comput. Commun. Control, 5(4), pp. 625-633. DOI: 10.15837/ijccc.2010.5.2218.

Conference Articles

- **Braşoveanu, A.M.P.**, Weichselbraun, A., Nixon, L.J.B. (2020). In Media Res: A Corpus for Evaluating Named Entity Linking with Creative Works. In Proceedings of the 24th Conference on Computational Natural Language Learning (pp. 355-364). DOI: 10.18653/v1/2020.conll-1.28
- Weichselbraun, A., **Braşoveanu, A.M.P.**, Waldvogel, R., Odoni, F. (2020). Harvest - An Open Source Toolkit for Extracting Posts and Post Metadata from Web Forums. IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology: 430-436.
- **Braşoveanu, A.M.P.**, Andonie, R. (2020) Visualizing Transformers for NLP: A Brief Survey. IV 2020: 270-279. DOI: 10.1109/IV51561.2020.00051.
- Weichselbraun, A., **Braşoveanu, A.M.P.**, Kuntschik, P., Nixon, L.J.B. (2019). Improving Named Entity Linking Corpora Quality. RANLP 2019, Varna, Bulgaria. Incoma. Published by ACL. pp. 1329-1338. DOI: 10.26615/978-954-452-056-4_152.
- Weichselbraun, A., Kuntschik, P., **Braşoveanu, A.M.P.** (2019). Name Variants for Improving Entity Discovery and Linking. LDK 2019. OASICS 70, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik 2019, pp. 14:1-14:15. DOI: 10.4230/OASIcs.LDK.2019.14.
- **Braşoveanu, A.M.P.,** Andonie, R. (2019). Semantic Fake News Detection: A Machine Learning Perspective. IWANN 2019 Part I, Springer, pp. 656-667. DOI: 10.1007/978-3-030-20521-8_54.
- Odoni, F., **Braşoveanu, A.M.P.**, Kuntschik, P., Weichselbraun, A. (2019). Introducing orbis: An extendable evaluation pipeline for named entity linking performance drill-down analyses. Proceedings of the Association for Information Science and Technology, 56(1), 468-471. DOI: 10.1002/pra2.49.
- Odoni, F., Kuntschik, P., **Braşoveanu, A.M.P.**, Weichselbraun, A. (2018). On the Importance of Drill-Down Analysis for Assessing Gold Standards and Named Entity Linking Performance. Semantics 2018, Procedia Computer Science, 137, 33-42. Elsevier. DOI: 10.1016/j.procs.2018.09.004.
- **Brasoveanu, A.M.P.**, Rizzo, G., Kuntschik, P., Weichselbraun, A., Nixon, L.J.B. (2018). Framing Named Entity Linking Error Types. LREC 2018, pp. 266-271. ELRA, Paris, France. DOI: www.lrec-conf.org/proceedings/lrec2018/pdf/612.pdf
- Weichselbraun, A., Kuntschik, P., **Braşoveanu, A.M.P.** (2018). Mining and Leveraging Background Knowledge for Improving Named Entity Linking. WIMS 2018, pp. 27:1-27:11. ACM. DOI: 10.1145/3227609.3227670.

- **Braşoveanu, A.M.P.**, Nixon, L.J.B., Weichselbraun, A. (2018). Storylens: A multiple views corpus for location and event detection. WIMS 2018, pp. 30:1-30:4. ACM. DOI: 10.1145/3227609.3227674.
- Marx, E., Sherkarpour, S., Soru, T., **Braşoveanu, A.M.P.**, Saleem, M., Baron, C., Weichselbraun, A., Lehmann, J.,. Ngonga Ngomo, A.-C., Auer, S. (2017). Torpedo: Improving the State-of-the-Art RDF Dataset Slicing. ICSC 2017, pp. 149-156. IEEE.DOI: 10.1109/ICSC.2017.79.
- **Braşoveanu, A.M.P.**, Nixon, L.J.B. Weichselbraun, A., & Scharl, A. (2017). A Regional News Corpora for Contextualized Entity Discovery and Linking. LREC 2016, pp. 3333-3338. ELRA, Paris, France. DOI: www.lrec-conf.org/proceedings/lrec2016/summaries/835.html.
- Sabou, M., **Braşoveanu, A.M.P.**, Onder, I. (2015). Linked Data for Cross-Domain Decision-Making in Tourism. ENTER 2015, pp. 197-210. DOI: 10.1007/978-3-319-14343-9_15.
- **Braşoveanu, A.M.P.**, Hubmann-Haidvogel, A., Scharl, A. (2012). Interactive visualization of emerging topics in multiple social media streams. AVI 2012, pp. 530-533. DOI: 10.1145/2254556.2254655.
- Sabou, M., **Braşoveanu, A.M.P.**, Arsal, I. (2012) Supporting tourism decision-making with linked data. I-SEMANTICS 2012, pp. 201-204. DOI: 10.1145/2362499.2362533.
- Hubmann-Haidvogel, A., **Braşoveanu, A.M.P.**, Scharl, A., Sabou, M., Gindl, S. (2012). Visualizing Contextual and Dynamic Features of Micropost Streams. MSM 2012, pp. 34-40. DOI: http://ceur-ws.org/Vol-838/paper_05.pdf.
- Oprean, C., Kifor, C., Barbat, B.E., **Braşoveanu, A.M.P.**, Fabian, R.D. (2010). Bounded Rationality in Computer Science Curricula. FECS 2010, pp. 135-140.