



HAL
open science

Evaluarea și sumarizarea automată a conversațiilor chat

Mihai Dascălu, Ștefan Trăușan-Matu, Philippe Dessus

► To cite this version:

Mihai Dascălu, Ștefan Trăușan-Matu, Philippe Dessus. Evaluarea și sumarizarea automată a conversațiilor chat. Romanian Journal of Human - Computer Interaction, 2010, Proceedings of the National Conference on Human-Computer Interaction Bucharest, 2-3 September 2010, 3, pp.95-100. hal-01079347

HAL Id: hal-01079347

<https://hal.science/hal-01079347v1>

Submitted on 1 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluarea și sumarizarea automată a conversațiilor chat

Mihai Dascălu

Universitatea “Politehnica” din
București, SC CCT SRL
313 Splaiul Independenței,
060042 București, România
mihai.dascalu@cs.pub.ro

Ștefan Trăușan-Matu

Universitatea “Politehnica” din
București
313 Splaiul Independenței,
060042 București, România
stefan.trausan@cs.pub.ro

Philippe Dessus

Grenoble University
1251 av. Centrale, BP 47, F-
38040 Grenoble CEDEX 9,
France
philippe.dessus@upmf-
grenoble.fr

REZUMAT

Odată cu evoluția continuă a mediilor colaborative, nevoia unei analize automate în vederea evaluării participanților la discuții de tip Instant Messenger (chat) a devenit din ce în ce mai acută. De asemenea, o facilitare necesară la nivelul unui sistem de notare și evaluare a replicilor și, implicit, a participanților este sumarizarea. În vederea atingerii acestor obiective, o serie de factori bazați pe procesarea limbajului natural (inclusiv analiza semantică prin intermediul *Latent Semantic Analysis* - LSA) și data-mining au fost luați în considerare. De asemenea, în vederea obținerii unei evaluări minuțioase și multi-perspectiviste a participanților, au fost calculați factori derivați din metodele lui Page de notare a eseurilor, diverse abordări pentru estimarea lizibilității și metrici specifice analizei rețelelor sociale. Un standard de aur a fost utilizat pentru măsurarea corectitudinii rezultatelor obținute.

Cuvinte cheie

sumarizarea extractivă a chat-urilor, evaluarea participanților din mediul colaborativ, rețele sociale, LSA etichetat.

Clasificare ACM

H5.2. Information interfaces and presentation: Training, help, and documentation and Natural language.

INTRODUCERE

Ca urmare a evoluției web-ului au apărut noi instrumente de colaborare și concomitent cu acestea s-a multiplicat dorința de procesare automată a unui volum cât mai mare de informații. În cadrul Învățământului Colaborativ Sprijinit de Calculator (*Computer Supported Collaborative Learning* – CSCL) [8], chat-urile joacă un rol important și au devenit din ce în ce mai utilizate. Pe de altă parte, evaluarea manuală a unui chat este un proces consumator de timp pentru profesor, necesitând pentru analiză un timp mai mare decât durata efectivă a discuției. Astfel, necesitatea de a dezvolta aplicații care pot ajuta procesul de evaluare a devenit esențială. Din această perspectivă au fost vizate două direcții majore:

- Un sistem automat pentru evaluarea fiecărui participant la o sesiune de chat prin utilizarea unor metode specifice prelucrării limbajului natural, analiza de suprafață și analiza rețelelor sociale;
- Un sistem automat de sumarizare extractivă bazat pe rezultatele sistemului de mai sus, care poate fi utilizat pentru a reduce considerabil timpul petrecut de tutor

în cazul unei evaluări manuale. Utilitatea sumarizării se răsfrânge în posibilitatea de a evalua manual un număr mult mai mare de chat-uri în aceeași unitate de timp. De asemenea, sumarizarea este utilă și unui participant care intră mai târziu în chat sau pentru a revedea ce s-a discutat.

Scenariul utilizat pentru sesiunile de chat a fost următorul: studenții din cadrul cursurilor de Interacțiune Om-Calculator (IOC) din doi ani diferiți și de la Prelucrarea Limbajului Natural, de la facultatea de Automatică și Calculatoare au fost invitați să discute beneficiile instrumentelor colaborative on-line (de exemplu, la IOC wiki-uri, blog-uri, chat-uri și forumuri), dezbătând și argumentând beneficiile și dezavantajele fiecărei tehnologii. Fiecare student era responsabil de o tehnologie și își susținea punctul de vedere relativ la beneficiile acesteia, încercând totodată să contracareze “oferțele” celorlalți participanți. În ultima parte a discuției erau analizate posibilitățile de integrare și inter-comunicare a diferitelor tehnologii. Toate interacțiunile sunt exclusiv între studenți, fără intervenția efectivă a unui profesor, iar limba utilizată în cadrul discuțiilor a fost engleza. Mediul utilizat a fost VMT [9], iar în medie fiecare chat a avut 4 participanți. Corpusul obținut conținea aproximativ 80 de chat-uri, care au fost ulterior evaluate manual de un student de la un an diferit. Astfel s-a încercat diminuarea subiectivismului în evaluare. Domeniul discuțiilor a fost IT-ul, centrat pe tehnologii web colaborative, iar participanții au avut cunoștințe similare, atât de bază cât și în domeniul tehnologiilor discutate, și aceleași roluri în cadrul discuției din perspectiva responsabilităților. Această analiză poate fi extrapolată la orice tip de discuții centrate pe un domeniu specific, cu o reprezentare corespunzătoare a cunoștințelor (condiționat, de un proces anterior de învățare supervizată folosind *Latent Semantic Analysis* (LSA)). De asemenea, se pot realiza experimente pe baza următoarelor scenarii: prezența efectivă a unui tutore/profesor și influența acestuia, sau interacțiunile între participanți la același nivel, precum este cazul celor prezentate anterior.

Următoarea secțiune a acestui document va prezenta factorii utilizați în cadrul procesului de evaluare: de la cei mai simpli precum lizibilitatea sau metricile derivate din metodele lui Page de analiză a eseurilor [7], trecând la analiza rețelelor sociale și, în final va fi utilizat *Latent Semantic Analysis* LSA [12], pentru o abordare semantică a procesului de notare. A treia secțiune evaluează sistemul, iar ultima secțiune, înainte de concluzii, va fi dedicată prezentării sistemului de sumarizare.

PROCESUL DE EVALUARE

Noutatea abordării constă în primul rând în procesul de evaluare atât al replicilor, cât și al participanților la nivelul căruia sunt integrate perspective multiple pentru o privire cât mai holistică. Astfel, evaluarea se realizează pe 3 nivele după cum urmează:

- Analiza de suprafață structurată pe două direcții: pe de-o parte evidențierea lizibilității replicilor, iar pe de alta o serie de metrice derivate din studiile inițiale ale lui Page pentru analiza eseurilor precum fluență, corectitudine gramaticală, dicție sau structura replicilor [7];
- Analiza morfologică (Spellchecking, Stemming, Tokenizing) și etichetarea actelor de vorbire (*Part of Speech Tagging*);
- Analiză semantică având în spate modelul Latent Semantic Analysis (LSA) pentru evaluarea similarității dintre o replică și întreaga conversație, respectiv cu un set predefinit de cuvinte specificate de profesor drept subiecte impuse de discuție.

Adițional, analiza specifică rețelelor sociale este aplicată atât asupra grafului rezultat din numărul de mesaje interschimbate între participanți, cât și asupra grafului cu costul arcelor egal cu suma notelor replicilor aferente.

Analiza de suprafață

Analiza de suprafață se focalizează pe două categorii de factori: lizibilitate și factori obținuți prin adaptarea studiilor lui Page aplicate inițial exclusiv pentru analiza eseurilor. Ideea în sine era simplă – evaluarea automată a unui document cu performanțe similare cu ale unui evaluator uman având la bază exclusiv atribute simple statistice [11]. Scopul sistemului presupunea facilitarea evaluării, permițând astfel asignarea mai multor teme și îmbunătățirea capacităților studentului prin exercițiu, similare cu cele ale sistemului de față. Ideea de evaluare a avut la bază relația dintre *proxes* (aproximări ale interesului documentului la nivel de factori ușor de calculat automat – practic componenta computațională) și *trins* (variabile intrinseci utilizate de către evaluatorii umani). Concluziile studiului inițial au fost promițătoare: corelație de 71% similară cu corelația medie dintre doi evaluatori umani, demonstrând astfel viabilitatea metodei.

Pornind de la acest studiu inițial și integrând abordarea lui Slotnick [11] pentru gruparea factorilor identificați anterior în funcție de valorile intrinseci, următorii factori au fost obținuți, fiecare cu o pondere egală la nivelul notei categoriei din care fac parte. Nota finală reprezintă media aritmetică a notele fiecărei categorii. Astfel factorii utilizați în special din perspectiva unei analize cantitative, ulterior convertiți în procente prin scalarea raportat la valoarea maximă, sunt cei din Tabelul 1 și sunt figurați conform Figurii 1 la nivelul aplicației dezvoltate.

A doua categorie de factori reprezentativi la nivelul de suprafață reflectă lizibilitatea textului înțeleasă drept ușurința cu care pot fi parcurse replicile unui participant, oferind o imagine concludentă asupra stilului propriu de scriere. Aceasta are un impact major asupra înțelegerii, retenției, vitezei de citire, precum și asupra persistenței textului citit.

Tabelul 1. Factori specifici analizei de suprafață

Factor	Proxes aferente
Fluență	Număr total de caractere, numărul total de cuvinte, numărul total de replici, numărul mediu de cuvinte per replică, numărul de cuvinte diferite utilizate, numărul de fraze (diferit de numărul de replici, întrucât se pot întâlni mai multe fraze la nivelul aceleiași replici)
Corectitudine gramaticală	Număr de cuvinte scrise corect din perspectivă gramaticală raportat la numărul total de cuvinte redactate de un utilizator
Dicție	Valoarea medie și abaterea standard relativ la lungimea cuvintelor utilizate
Structura replicilor	Numărul de replici, numărul mediu de cuvinte, lungimea media a unei replici estimată prin numărul de caractere

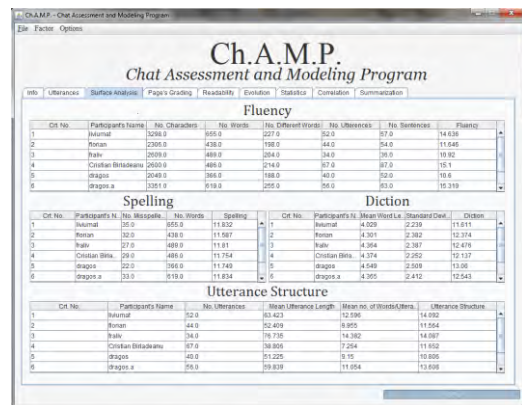


Figura 1. Analiza de Suprafață – metrice având la bază Page's proxes

Un alt aspect important relativ la lizibilitate într-un mediu colaborativ îl reprezintă reflectarea punctului de vedere al participantului din perspectiva trăsăturilor personale, interes, motivare sau chiar cunoștințe apriori. Un limbaj elevat induce o anumită stare, pe când utilizarea permanentă a unor replici scurte creează o cu totul altă impresie.

Întrucât discuțiile sunt purtate în Engleză și pornind de la citatul lui Jacques Barzun conform căruia „*Simple English is no person's native tongue*”, adresarea corespunzătoare audienței, în cazul nostru formată din ceilalți participanți la chat, este esențială. În cadrul discuțiilor purtate există o simplificare a frazelor; totuși, lizibilitatea oferă în continuare informații utile despre nivelul de instruire, de înțelegere și uneori atitudine. Oricum corelarea cu alte metrice, descrise ulterior, joacă un rol esențial.

În mod uzual, lizibilitatea este estimată în mod inconștient, însă din perspectivă computațională au fost folosite următoarele formule obținute în cadrul unor studii detaliate, pe un eșantion mare de texte și care aproximează nivelul țintă al audienței:

- *Flesch Reading Ease Readability Formula* [13] este una din cele mai vechi abordări care exprimă dificultatea citirii unui text pe o scală de 0 la 100. O valoare mai mare exprimă o ușurință mai mare de citire, nu neapărat și o înțelegere mai bună a textului. Plaja optimă este considerată [60, 70]:

$$RE = 206,835 - (1,015 * ASL) - (84,6 * ASW)$$

RE = Readability Ease

ASL = Average Sentence Length (numărul de cuvinte raportat la numărul de fraze)

ASW = Average number of Syllables per Word (numărul total de silabe împărțit la numărul de cuvinte)

- *Gunning Fog Index (FOG)* [14] pleacă de la ideea că articolele din ziar și documentele business sunt pline de ambiguități și informații inutile ("fog") având astfel o complexitate inutilă. Astfel, formula estimează numărul de ani de educație formală pe care un cititor de inteligență medie ar trebui să-i aibă în spate pentru a înțelege un text la nivelul primei parcurgeri. Un dezavantaj al implementării este că nu toate cuvintele de peste 2 silabe sunt dificile, dar performanțele computaționale sunt foarte bune. Valorile optime sunt în jurul lui 12, iar orice peste 18 este mult prea dificil de citit:

$$FOG = (ASL + PHW) * 0,4$$

PHW = Percentage of Hard Words (în implementarea curentă numărul de cuvinte cu mai mult de 2 silabe relativ la mulțimea tuturor cuvintelor utilizate)

- *Flesch Grade Level Readability Formula* [15] notează participanții după modelul de școlarizare american. Astfel se formează o asociere biunivocă: scorul N determină faptul că documentul poate fi citit de o persoană cel puțin în clasa a N-a. Abordarea ajută și la asignarea de materiale de citit conform nivelului actual de instruire. O valoare în domeniul [7, 8] este considerată optimă întrucât textul este ușor de citit:

$$FKRA = (0,39 * ASL) + (11,8 * ASW) - 15,59$$

FKRA = Flesch-Kincaid Reading Age

Notarea participanților din cadrul unui chat se face în funcție de abaterea față de valorile considerate optime, iar vizualizarea se realizează conform următoarei figuri.

Ch. No.	Participant's Name	Flesch score	Reading Ease
1	Ionela	74,500	Difficult
2	Stelian Brabantu	75,251	Difficult
3	Renan	76,443	Difficult
4	Andi	88,463	Standard - optimal
5	Dragos A	89,074	Standard - optimal
6	Dragos	87,243	Standard - optimal

Ch. No.	Participant's Name	Years of formal education	Fog complexity
1	Ionela	8,932	Complex
2	Dragos	8,498	Complex
3	Dragos A	8,712	Complex
4	Andi	8,661	Complex
5	Stelian Brabantu	8,402	Complex
6	Ionela	8,261	Complex

Ch. No.	Participant's Name	Score	U.S. grade school level
1	Andi	7,854	8th grade - optimal
2	Stelian Brabantu	8,294	8th grade
3	Andi	8,159	8th grade
4	Dragos A	8,395	8th grade
5	Ionela	8,718	8th grade
6	Dragos	8,628	8th grade

Figura 2. Lizibilitatea

Analiza rețelelor sociale

În scopul obținerii unei perspective cât mai profunde, factorii sociali sunt luați în considerare pentru reliefaarea interacțiunii dintre participanți. Astfel, un graf este generat pe baza transcriptului chat-ului analizat: nodurile reprezintă participanții, iar arcele sunt generate pe baza legăturilor explicite prin posibilitățile de referențiere directă între replici din mediul chat.

Din punctul de vedere al rețelelor sociale, diverse metrice sunt calculate pentru a determina cel mai implicat participant la nivelul discuției: gradul (*indegree* și *outdegree*), centralitate (*closeness*, *graph centrality*, *eigen centrality*) și *ranking*-ul utilizatorilor, similar cu binecunoscutul algoritm Google Page Rank. Aceste metrice sunt aplicate asupra unor matrice la nivelul cărora arcele reprezintă:

- Numărul efectiv de replici interschimbate între participanți, cu accent preponderent pe o abordare cantitativă;
- Suma notelor fiecărei replici bazate pe evaluarea semantică a fiecărei replici; procesul de evaluare propriu-zisă a unei replici va fi descris într-o secțiune ulterioară, iar această abordare fundamentează evaluarea calitativă a chat-ului.

Astfel, din aplicarea metodelor specifice de analiză a rețelelor sociale, atât asupra numărului de replici, cât și asupra sumei notelor aferente, se accentuează dualitatea abordării din următoarele perspective:

- Evaluare cantitativă vs. Evaluare calitativă cu substrat semantic;
- Implicarea în discuție (*gregariousness*) vs. cunoștințele prezentate/dobândite și aportul cognitiv, evaluate semantic folosind LSA.

Toate metricile folosite în cadrul analizei rețelelor sociale reprezintă valori *relative* în sensul în care pot oferi scoruri relevante numai în comparație cu implicarea celorlalți participanți ai aceluiași chat. Acesta este principalul motiv pentru care toți factorii sunt scalați între participanți, oferind fiecăruia un procent ponderat raportat la performanța globală.

Un rol important în analiza rețelelor sociale îl are componenta de vizualizare, care permite modelarea atât prin intermediul unui model fizic, cât și prin intermediul unui model radial al interacțiunii dintre participanți [2].

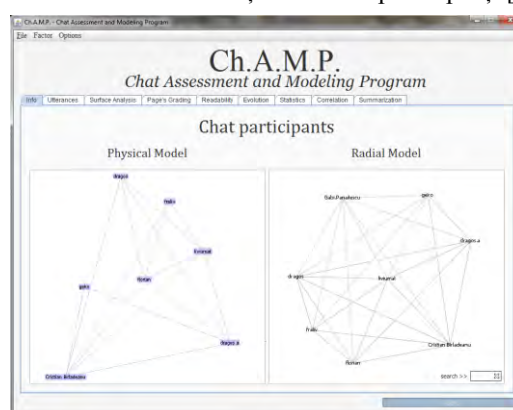


Figura 3. Modelarea rețelelor sociale

Analiza semantică – LSA

Analiza Semantică Latentă este o tehnică ce se bazează pe un *model al spațiului de vectori* care este folosit în analiza similarității dintre un set de documente și termenii conținuți. LSA pornește de la o matrice termen-document bi-dimensională care descrie numărul de apariții ale fiecărui termen în documentul corespunzător. LSA transformă matricea inițială într-o relație dintre termeni și concepte, respectiv o relație între aceste concepte și

documentele aferente. Astfel, termenii și documentele devin în mod indirect legate prin concepte [4, 6]. Această transformare se obține printr-o descompunere în valorile singulare (*Singular Value Decomposition* SVD) și, ulterior, printr-o reducere a dimensionalității prin proiecție după k stabilit.

Procesul de învățare supervizată folosește corpul de chat-uri. Primul pas în procesul de învățare, după verificarea ortografiei, este de a elimina cuvintele cu aport semantic redus – *stop-words* (foarte frecvente și lipsite ca importanță globală "the", "a", "an", "to", etc.) din fiecare replică. Următorul pas este etichetarea actelor de vorbire – *Part of Speech Tagging* – iar, în cazul verbelor, acestea sunt reduse la rădăcină prin *stemming* pentru limitarea formelor identificate, toate exprimând același concept. Toate celelalte cuvinte sunt lăsate în formele lor identificate, singura îmbunătățire existând prin adăugarea de etichete corespunzătoare actelor de vorbire. Se consideră că două cuvinte, identice ca rădăcină și formă, dar cu acte de vorbe diferite se regăsesc în contexte diferite, prin urmare învecinându-se cu alte noduri din perspectivă semantică [10].

După popularea matricei termen-document se aplică *term frequency - inverse document frequency* - TF-IDF (dependență directă relativ la frecvența termenilor și inversă raportat la numărul de documente [6]). Etapele finale ale procesului de antrenare sunt descompunerea în valori singulare (SVD) și proiecția matricelor obținute cu scopul de a reduce dimensiunile acestora. În conformitate cu [5], valoarea optimă empirică pentru k este 300, o valoare utilizată frecvent în experimente curente și asupra căreia mai multe surse sunt de acord.

Un alt aspect important în procesul de învățare LSA este segmentarea care presupune împărțirea unui chat luând în considerare unități semnificative. În implementarea actuală chat-ul este împărțit între participanți, din cauza unității și coeziunii între replicile aceleiași persoane. Aceste documente sunt ulterior împărțite în segmente folosind ferestre disjuncte, fixe ca dimensiune.

Scopul final al utilizării LSA este de a evalua proximitatea/similaritatea dintre doi termeni sau cuvinte utilizând drept măsură cosinusul dintre vectorii aferenți:

$$Sim(cuv_1, cuv_2) = \frac{\sum_{i=1}^k cuv_{1,i} \cdot cuv_{2,i}}{\sqrt{\sum_{i=1}^k cuv_{1,i}^2} \times \sqrt{\sum_{i=1}^k cuv_{2,i}^2}}$$

Similaritatea dintre replici și similaritatea dintre replici și întregul document sunt utilizate pentru a evalua importanța fiecărei replici în parte, comparativ cu întreaga discuție sau cu un set predefinit de cuvinte cheie menționate ca un nou document:

$$Vector(repl) = \sum_{i=1} (1 + \log(nr_aparitii(cuv_i))) * vector(cuv_i)$$

$$Sim(repl_1, repl_2) = Sim\left(\begin{matrix} Vector(repl_1) \\ Vector(repl_2) \end{matrix}\right)$$

Mecanismul de notare al replicilor, respectiv al participanților

Primul aspect care trebuie luat în considerare îl reprezintă construirea grafului bazat pe legături explicite pentru marcarea relațiilor existente. Procesul de evaluare al fiecărei replici se bazează pe textul prelucrat și obținut în urma analizei morfologice și a *POS Tagging*-ului și presupune următoarele etape:

1. *Evaluarea fiecărei replici individual*, luând în considerare următoarele caracteristici:

- lungimea efectivă a replicii inițiale;
- lista de cuvinte rămase după eliminarea cuvinte de tip *stop-words*, verificate în prealabil ortografic; de asemenea se memorează și numărul lor de apariții;
- nivelul pe care se află replica actuală într-un fir al discuției;
- factorul de ramificare egal cu numărul de replici care au legături explicite către replica curentă;
- similaritatea cu întreaga discuție pentru a evidenția importanța replicii raportat la întreaga conversație;
- similaritatea cu un set predefinit de subiecte de discuție asigurate de profesor pentru a verifica acoperirea temei date pentru discuție.

Aspectul surprins în cadrul abordării este exact dualitatea perspectivei: pe de-o parte avem abordarea cantitativă (lungimea replicii exprimată în caractere scrise, pornind de la premisa că informația ar trebui să fie mai valoroasă dacă este transmisă în mai multe mesaje, legate împreună, și o lungime mai mare nu este datorată dorinței de a impresiona, ci pentru îmbogățirea contextului), iar de cealaltă parte abordarea calitativă care presupune utilizarea LSA și a cuvintelor cheie predefinite.

Formule utilizate pentru evaluarea fiecărei replici sunt:

$$SCOR_{empiric} = \left(\frac{\text{lungime}(repl_neprelucrata)}{10} + \frac{9}{10} \times \sum_{word}^{remaining} scor(cuv) \right) \times factor_augmentare$$

$$\begin{aligned} scor(cuv) &= \text{lungime}(radacina) * (1 + \log(nr_aparitii)) \\ augmentare &= (1 + \log(nivel)) \times (1 + \log(factor_ram)) \times \\ &\quad \times Sim(replica, conversatie) \times \\ &\quad \times Sim(replica, set_subiecte_discutie) \end{aligned}$$

2. *Augmentarea scorului*

Fiecare fir de discuție are un maxim global relativ la scorurile calculate anterior în jurul căruia toate replicile au importanța augmentată corespunzător cu o distribuție Gaussiană:

$$\begin{aligned} p(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ unde:} \\ \sigma &= \frac{\max(id_repl_fir) - \min(id_repl_fir)}{2}, \\ \mu &= id_replica_cu_scor_max_im \end{aligned}$$

Astfel scorul fiecărei replici este amplificat cu un factor de $1 + p(replica_curenta)$.

3. Determinarea scorului final de-a lungul firului de discuție

Pe baza scorului empiric, scorul final este obținut după următoarea formulă:

$$scor_{final} = scor_{final}(repl_precedenta) + coef \times scor_{empiric},$$

unde coeficientul este determinat de tipul actelor de vorbire din replica curentă și din cea de care aceasta este legată printr-o legătură explicită; astfel, sunt inspectate verbele, semnele de punctuație și anumite cuvinte cheie. Pornind de la un set de tipuri de acte de vorbire, coeficienții sunt obținuți dintr-o matrice predefinită; valorile au fost determinate după estimarea impactului replicii curente luând în considerare numai nivelul anterior de replici din cadrul aceluiași fir de discuție (similar unui proces Markov). Scorul final al replicilor din cadrul unui fir de discuție poate fi ridicat sau coborât de fiecare replică. Prin urmare, în funcție de tipul replicii și al actelor de vorbire identificate, nota finală poate avea o valoare atât pozitivă cât și negativă.

Notarea participanților combină metricile de la nivelul analizei de suprafață cu factorii specifici analizei rețelelor sociale aplicați asupra celor 2 matrice definite anterior. Fiecare dintre factorii de analiza este convertit la un procent (valoare curentă / suma tuturor valorilor, pentru fiecare factor, cu excepția cazului de *eigen-centrality* când conversia se face automat prin înmulțirea cu 100 a valorii proprii în modul). Scorul final al participanților ia în considerare toți acești factori și ponderile corespunzătoare:

$$scor_final_i = \sum_k pondere_k \times procent_{k,i},$$

unde k este un factor al participantului cu rangul i .

Corelația Pearson pentru fiecare factor este determinată în raport cu notele oferite de evaluatori umani, permițând astfel evaluarea importanței și a relevanței fiecăruia relativ la notele date de un evaluator uman luate drept standard de referință.

EVALUAREA SISTEMULUI

Ponderile asociate static fiecărui factor sunt următoarele: 10% pentru metricile derivate din metoda lui Page de evaluare a eseurilor și lizibilitate, 5% pentru factorii de analiză a rețelelor sociale aplicați pe numărul de replici interschimbate și 10% pentru factorii semantici din analiza rețelei sociale construită pe baza scorurilor fiecărei replici.

Rezultatele globale obținute cu aceste ponderi pentru corpusul analizat sunt:

- corectitudinea relativă: 77.44%
- corectitudine absolută: 70.07%
- corelația Pearson: 51,4%

Corectitudinea relativă și corectitudinea absolută reprezintă distanțe absolute/relative dintre valorile obținute automat și cele din urma evaluării umane. Rezultatele finale (ca media aritmetică pentru fiecare din cei 3 factori determinați individual per chat) sunt afișate la nivelul interfeței grafice proiectate:

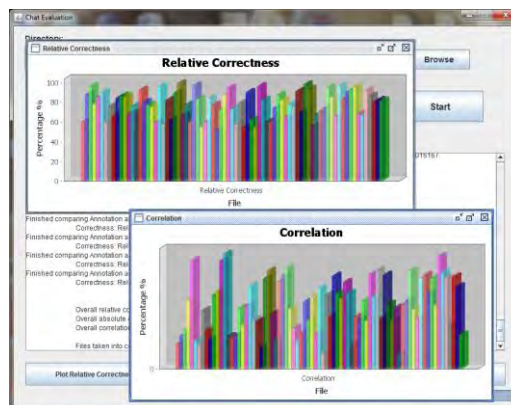


Figura 4. Rezultatele obținute

SUMARIZARE EXTRACTIVĂ

Cel mai important aspect relativ la generarea extractivă a unui rezumat este legătura strânsă dintre sumarizare și evaluarea propriu-zisă a chat-ului. Prin urmare, pentru a obține un rezumat extractiv cât mai corect, fiecare replică trebuie evaluată - în cazul nostru, procesul de notare a fost descris anterior și oferă baza pentru identificarea replicilor cheie. Aceasta abordare combină metoda de sumarizare automată bazată pe LSA [3], cu beneficiile de a introduce toți ceilalți factori prezentați în secțiunile precedente, cu impact mare în procesul de evaluare.

Pentru a răspunde cât mai bine cerințelor utilizatorului, un set de procente predefinite (procente incrementale cu 10%) a fost prevăzut, permițând utilizatorului să selecteze nivelul de interes și de filtrare al replicilor. Astfel se poate determina numărul de replici care urmează să fie selectate și care vor face parte din rezumatul oferit de către sistem. Ulterior, replicile sunt extrase în ordinea descrescătoare a importanței / scorului. Pentru a augmenta rolul replicilor direct legate de replica curentă (implicit corelate și din perspectiva discursului) se acordă o bonificație tuturor replicilor de pe următorul nivel din cadrul arborelui de replici definit pe baza legăturilor explicite. Astfel se asigură posibilitatea selectării comentariilor și impresiilor referitoare la o anumită replică, chiar dacă scorul inițial al acestora este relativ mic. Într-o sumarizare a chat-urilor este esențială integrarea acestor replici datorită aportului cognitiv și importanței globale relativ la impresiile împărtășite.

Scopul final al sistemului este de a reduce drastic timpul petrecut de un evaluator uman pentru a nota un chat, prin urmare, să permită relocarea timpului înspre sarcini mai creative, fără a elimina componenta umană în evaluare.

Impactul real în urma utilizării sistemului automat de rezumare este reducerea cu mai mult de 30% a timpului petrecut pentru evaluarea unui singur chat. Acest lucru permite analiza mai multor chat-uri, prin urmare oferind un corpus adnotat mai mare, util în procesul de învățare automată. Astfel, rezultatele pot fi îmbunătățite prin ajustarea fină a ponderilor pentru fiecare dintre factorii luați în considerare în procesul de notare.

Acest rezumat este, de asemenea, foarte util din perspectiva studentului, făcând posibilă o scurtă descriere a fiecărei discuții, evidențierea enunțurilor cu un interes mai mare și bogate în informații.

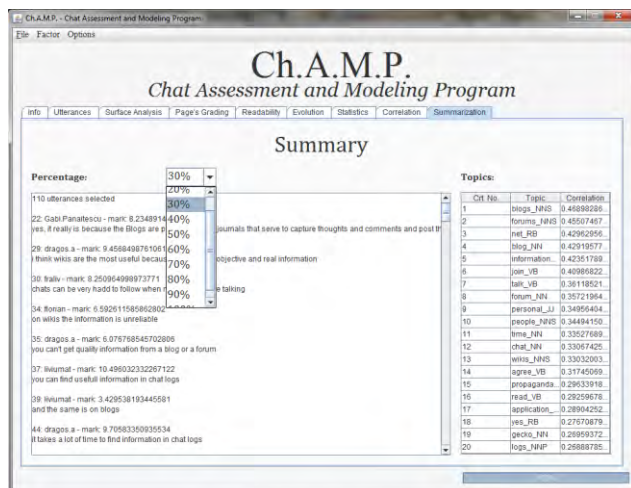


Figura 5. Sumarizare automată

CONCLUZII

Rezultatele obținute ne permit să concluzionăm că evaluarea contribuției participanților într-un mediu colaborativ poate fi realizată automat, iar impactul la nivelul tutorelui se reflectă într-o scădere cu peste 30% a timpului petrecut pentru evaluarea unui chat. De asemenea, împreună cu o ajustare mai bună a parametrilor și a ponderilor aferente fiecărui factor, împreună cu un proces îmbunătățit de învățare supervizată utilizând LSA, precum și prin creșterea numărului de factori specifici analizei rețelelor sociale (inclusiv cei aplicați la nivelul întregii rețele) credem că performanța și fiabilitatea rezultatelor obținute va crește. Mai mult, factorul subiectiv în evaluarea manuală este întotdeauna prezent și influențează corectitudinea per ansamblu a sistemului.

Pornind de la direcțiile descrise anterior, următoarele îmbunătățiri vor fi luate în considerare pentru îmbunătățirea rezultatelor:

- experimente sociale cu evaluarea impactului aferent; realizarea de scenarii diferite (cu sau fără prezența efectivă a unui profesor) corelate cu impactul efectiv al prezenței acestuia;
- identificarea de artefacte, voci și perspective multiple ale aceluiași concept folosind LSA, în concordanță cu filosofia lui Bakhtin [1] pentru o înțelegere mai profundă a discuției din punct de vedere dialogistic;
- includerea de texte comune (de exemplu, articole de ziar), pentru o perspectivă de ansamblu asupra tuturor cuvintelor întâlnite dintr-un text (nu numai asupra celor specifice unui anumit domeniu) și pentru a îmbunătăți procesul de învățare LSA;
- integrarea tuturor chat-urilor și generarea unei rețele sociale globale, rezultând astfel posibilitatea unei evaluări de ansamblu asupra întregului corpus și posibilitatea de a căuta un participant pe baza competențelor sale evaluate raportat la un subiect predefinit.

MULȚUMIRI

Ne dorim să mulțumim tuturor studenților din cadrul Facultății de Automatică și Calculatoare, Universitatea "Politehnica" din București, care au participat la experimentele noastre și care au format baza pentru corpusul utilizat pentru evaluarea sistemului. Activitățile de cercetare prezentate în această lucrare au fost parțial efectuate în cadrul proiectului PC7 LTFLL (Language Technologies for Lifelong Learning).

REFERINȚE

1. Bakhtin, M. M.: Problems of Dostoevsky's poetics (Edited and translated by Caryl Emerson). Minneapolis: University of Minnesota Press (1993)
2. Dascalu, M., Trausan-Matu, St., „Ch.A.M.P. – A Program for Chat Modeling and Assessment”, Annals of the Academy of Romanian Scientists Series on Science and Technology of Information Volume 2, Number 1/2009, ISSN 2066 – 2742, pag 95-106
3. Fernandez, S., Velazquez, P., Mandin, S.: Les systèmes de résumé automatique sont-ils vraiment des mauvais élèves?. In: JADT 2008: 9es Journées internationales d'Analyse Statistique des Données Textuelles (2008)
4. Landauer, K. Th., Foltz, W. P., Laham, D.: An Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284 (1998)
5. Lemaire, B.: Limites de la lemmatisation pour l'extraction de significations. JADT 2008: 9^{es} Journées internationales d'Analyse statistique des Données Textuelles (2008)
6. Manning, C., Schütze, H.: Foundations of statistical Natural Language Processing. MIT Press, Cambridge (Mass.) (1999)
7. Page, E. B. Paulus, D. H.: Analysis of essays by computer. Predicting Overall Quality, U.S. Department of Health, Education and Welfare (1968)
8. Stahl, G., Group Cognition: Computer Support for Building Collaborative Knowledge. MIT Press (2006) How to Classify Works Using ACM's Computing Classification System: <http://www.acm.org/class/>
9. Stahl, G. (Ed.), *Studying Virtual Math Teams* pag. 451 – 473, Boston, MA: Springer US
10. Wiemer-Hastings, P., Zipitria, I.: Rules for syntax, vectors for semantics. In: proceeding of the 23rd Annual Conference of the Cognitive Science Society (2001)
11. Wresch, W.: The Imminence of Grading Essays by Computer--25 Years Later. Computers and Composition 10(2), 45-58, retrieved from http://computersandcomposition.osu.edu/archives/v10/10_2_html/10_2_5_Wresch.html (1993)
12. <http://lsa.colorado.edu/>
13. <http://www.readabilityformulas.com/flesch-reading-ease-readability-formula.php>
14. <http://www.readabilityformulas.com/gunning-fog-readability-formula.php>
15. <http://www.readabilityformulas.com/flesch-grade-level-readability-formula.php>