



ACADEMIA ROMÂNĂ
SECȚIA DE ȘTIINȚA ȘI TEHNOLOGIA INFORMAȚIEI

INSTITUTUL DE CERCETARI PENTRU
INTELIGENȚĂ ARTIFICIALĂ

**Proiectarea învățării automate a extragerii de legături implicite
dintr-un corpus adnotat de conversații**

Ștefan Trăușan-Matu

Raport de Cercetare

BUCUREȘTI
Iunie 2009

CUPRINS

CUPRINS	1
INTRODUCERE	2
LEGĂTURILE IMPLICITE ȘI ÎNVĂȚAREA AUTOMATĂ A EXTRAGERII LOR	2
IMPORTANȚA ȘI CLASIFICAREA LEGĂTURILOR IMPLICITE	2
TIPURI DE LEGĂTURI IMPLICITE	4
DETECTAREA LEGĂTURILOR IMPLICITE PRIN ÎNVĂȚARE AUTOMATĂ.....	5
CODIFICĂRI ALE LEGĂTURILOR IMPLICITE ÎN ADNOTARE	7
SCHEMA DE ADNOTARE	15
EDITOR PENTRU ADNOTAREA CHATURILOR	19
CONCLUZII ȘI DIRECȚII VIITOARE DE CERCETARE.....	20
REFERINȚE	21

INTRODUCERE

Ca urmare a schimbărilor substanțiale introduse de Internet, Web și, în special, de instrumentele colaborative de interfațare specifice așa numitului "Web social", este normal să fie încercată folosirea acestor instrumente pentru îmbunătățirea proceselor de învățământ. Atracția unui astfel de scenariu este datorată și faptului că elevii și studenții petrec o parte din ce în ce mai mare în fața calculatorului, nu mai citesc aproape de loc cărți, manuale, în schimb explorează ore în șir păienjenişul WWW și interacționează o mare parte din timp folosind instrumente colaborative de gen mesagerie instantanee (sau, pe scurt "chat", de exemplu, Yahoo Messenger), emailuri, forumuri de discuții, bloguri sau rețele sociale.

Ca urmare a experiențelor nereușite de a înlocui total învățământul clasic, bazat pe prezența elevilor și a unui profesor într-o clasă fizică, cu un asistent automat, cu inteligență artificială [31], s-a propus un regim de "învățare hibridă" sau "combinată" ("blended", în engleză, vezi http://www.cs.cityu.edu.hk/~wbl2007/WBL2007_Proceedings_HTML/WBL2007_Proceedings.pdf). Învățământul colaborativ sprijinit de calculator (în engleză, "Computer Supported Collaborative Learning", CSCL) este o formă de învățământ hibrid care folosește noile instrumente, introduse în Webul social, pentru extinderea învățământului clasic. CSCL este însă și o nouă paradigmă care se impune pe fundamente teoretice cum ar fi construirea socială, în grup a cunoașterii [18].

Folosirea mesageriei instantanee (chat) în CSCL implică necesitatea existenței unor instrumente de analiză, validare și generare de feedback studenților care învață colaborativ [32]. O problemă foarte dificilă care trebuie rezolvată în acest context este identificarea firelor de discuție în conversațiile chat. Acestea sunt formate din succesiuni de legături explicite (date de student) sau implicite, care trebuie detectate de program. Acestea din urmă pot fi de diverse tipuri (coreferințe, perechi de adiacență, argumentări etc.) și pot fi detectate prin patternuri care pot fi definite de specialiști sau pot fi învățate automat. Această a doua variantă este subiectul prezentei lucrări.

LEGĂTURILE IMPLICITE ȘI ÎNVĂȚAREA AUTOMATĂ A EXTRAGERII LOR

Importanța și clasificarea legăturilor implicite

Analiza și eventuala prelucrare cu calculatorul a conversațiilor chat (de exemplu, efectuate în CSCL) include câteva clase importante de aplicații:

- rezumarea conversațiilor
- analiza contribuției fiecărui participant la o conversație
- analiza inter-animării, adică a procesului prin care participanții la discuție construiesc cunoaștere colaborativ folosind actele antevorbitorilor ("uptakes" [37]), care se constituie în transacte ("transacts", în engleză [38]).

Din punct de vedere teoretic, ne putem plasa în mai multe perspective, conform mai multor discipline (vezi și [39]): lingvistică (pragmatică, discurs), retorică, teoria argumentației [40], sociologie (analiza conversației [41]), antropologie lingvistică ([42]), literatură, filologie, dialogistică [3], psihologie, comunicare, filosofie [44] și semiotică. În multe din aceste discipline se consideră esențial rolul contextului, iar, spre deosebire de text, rostirea, intervenția ("utterance", în engleză) este unitatea de analiză, spre deosebire de propoziție (de exemplu, în lingvistică).

Spre deosebire de conversația "față în față" (f2f), în sistemele de mesagerie instantanee cu mai mulți participanți pot apare mai multe fire de discuție în paralel. Pentru a elimina ambiguitatea, sisteme cum ar fi ConcertChat [10] sau Polyphony [45] permit specificarea explicită a replicii la care se referă, la care se leagă replica curentă. De exemplu, în figura 1, coloana a doua conține aceste referințe și se pot remarca mai multe fire de discuție, ilustrate prin săgeți de stil diferit. Din experimentele făcute se constată că participanții recurg destul de des la fire paralele, în acest mod crescând volumul comunicării și, mai mult, apărând posibilitatea unor interferențe între subiectele paralele de discuție. În plus, în afara legăturilor explicite, firele de discuție generate folosesc curent legături implicite, care apar în orice text.

Nr	Ref	User	Text	Time
17		tim	You discussed about a topic separation	10.26.25
18	15	adrian	First of all, the reply method is cumbersome	10.26.37
19	17	john	yes.. because we did not like the way the topics were presented in concert chat	10.26.50
20	18	john	yes !!	10.26.56
21	20	john	i hate double-clicking !	10.27.04
22	20	tim	and how can we find topics ?	10.27.18
23	18	adrian	What bothers me is the linear presentation of the discussin	10.27.26
24	23	john	Yep	10.27.43
25	18	adrian	and double-clicking too	10.27.46
26		tim	You mean u want something like a chat forum ? :)	10.27.54
27	24	john	and the reply-to facility is supposed to help you	10.27.58
28	18	adrian	i'd like a tree presentation more	10.28.15
29	18	adrian	or maybe multiple chat columns, for each chat sub-thread	10.28.38
30	27	john	but it is really difficult to use in real-time, because there are so many topics discussed which intertwine each other	10.28.58
31	28	john	i subscribe to a tree-like presentation form	10.29.18
32	P 39	adrian	yes, that's why a clear separation of topics is needed	10.29.20
33	31	adrian	this is easy to implement, no problem here :)	10.29.47
34	30	tim	You need also a clever visual representation	10.29.49
35	30	tim	you'll need also a clever visual interface	10.30.05
36		tim	Who decides the topics ?	10.30.22
37	33	john	i suppose you are referring to the visual representation , right ?	10.30.33
38	33	john	What i would like is a clever way to separate the topics :)	10.30.45
39	38	john	not just doing ot myself, manually	10.30.59
40	37	adrian	Yeah	10.31.00
41	39	adrian	When you start a new thread (a new message, non-related to other message), the app can assume a new topic	10.31.44
42	39	john	i would like the application to be able to detect w topic change all by itself	10.31.46

Figura 1. Fire de discuție într-un chat

Tipuri de legături implicite

O clasificare a legăturilor implicite poate fi (vezi și [39]):

- repetiții de cuvinte, expresii
- colocații
- lanțuri lexicale,
- coreferințe,
- perechi de adiacență,
- acte de vorbire (în perechi de adiacență),
- implicaturi conversaționale (în sens Grice),
- între componentele structurilor retorice
- de argumentare (în sens Toulmin)
- transacte
- dialogistice (în sens Bahtin)

Un *lant lexical* este, potrivit definiției dată de Hirst și Budaniski [61], o succesiune de cuvinte înrudite semantic dintr-un text, care creează un anumit context și contribuie la continuitatea înțelesului. Bineînțeles că într-un text pot exista mai multe lanțuri lexicale în paralel. Datorită faptului că lanțurile lexicale sunt alcătuite din cuvinte înrudite semantic, ele pot fi calculate fără a fi nevoie de o înțelegere în profunzime a textului. Relațiile semantice dintre cuvinte nu sunt limitate în nici un fel de vreun anumit context al textului în care apar ci ele fac parte din gramatica limbii respective. Tocmai de aceea este posibil să calculăm lanțurile lexicale indiferent de contextul în care anumite cuvinte înrudite apar.

Teoria actelor de vorbire, introdusă de Austin și dezvoltată de Searle [61, p.137], pleacă de la ideea că există propoziții sau replici care se constituie în acte, despre care nu se poate spune că sunt adevărate sau false, ci important este efectul pe care îl au. Pentru analiza dialogurilor, teoria actelor de vorbire oferă un cadru de analiză a legăturilor implicite între replici.

Perechile de adiacență sunt compuse din două acte de vorbire successive care apar legate unul de altul, cum ar fi întrebare-răspuns, propunere-replică etc. Ideea perechilor de adiacență a fost introdusă în analiza conversației [41].

Teoria structurilor retorice (RST – Rhetorical Structure Theory) a fost introdusă de Mann și Thompson [43] și este una din cele mai răspândite teorii ale discursului în prelucrarea limbajului natural. În centrul acestei teorii stă ideea de relație retorică, care este o relație ierarhică între două unități de text, una fiind numită nucleu și cealaltă satelit. Există și excepții, relația de tip CONTRAST este o relație între două sau mai multe nuclee. O relație exprimă faptul că nucleul poate fi de sine stătător, iar satelitul ajută la înțelegerea nucleului. Empiric vorbind, nucleul exprimă ce este mai esențial în scopul scriitorului față de satelit și nucleul unei relații retorice este inteligibil independent de satelit, dar nu viceversa. Exemple de relații retorice sunt: ANTITEZĂ, SECVENȚĂ, JUSTIFICARE, CONCESIE, ELABORARE, CAUZĂ, EVIDENȚĂ, CONTRAST, EXEMPLU etc.

Prin *transact* se înțelege o acțiune conversațională prin care un participant operează asupra raționamentului, asupra ideilor altui participant. Se construiește astfel un discurs plecând de

la ideile partenerului de discuție prin continuare, extindere, punere la îndoială (dezbateri) sau integrare: "Participants in a collaborative learning setting are said to have a transactive discussion when they elaborate, build upon, question, or argue against the ideas presented by their partners in the process of working towards a common understanding of the task and reaching a shared solution." [53].

După cum remarcă Joshi și Rosé [53], spre deosebire de *Teoria Structurilor Retorice* (RST, [43]), în care relațiile informaționale conectează părți ale argumentării unui vorbitor, transactele sunt formulate pentru a reprezenta *relațiile între pozițiile în competiție ale mai multor vorbitori*. Tot ei remarcă faptul că, deși transactele sunt definite la același nivel ca *actele de vorbire*, care asignează o categorie semantică la contribuțiile la un dialog (cum ar fi întrebările și răspunsurile) granularitatea lor semantică diferă. De exemplu, un act de dialog de tip "întrebare" poate fi o cerere de feedback, o cerere de justificare sau o extensie. Diferențele între ele fiind importante pentru analiza succesului interacțiunii și al analizei înțelegerii.

Russel clasifică transactele după mai multe criterii. În primul rând, consideră că există replici transactive și replici netransactive. Apoi face o clasificare în întrebări, răspunsuri și afirmații ("statement"), iar, în funcție de destinatar, identifică transacte orientate către sine și transacte orientate către altul.

Detectarea legăturilor implicite prin învățare automată

În cadrul temei de față vom integra abordările "clasice" din prelucrarea limbajului natural (lanțuri lexicale, coreferințe, acte de vorbire, implicaturi conversaționale, discurs, teoria structurilor retorice), cu idei din dialogistica lui Bahtin [3] și analiza discursului conversațional [39].

În [33] a fost făcut un studiu în care au fost discutate în detaliu legăturile implicite specifice actelor de vorbire, perechilor de adiacență, teoriei structurilor retorice și lanțurilor lexicale. Tot acolo au fost amintite și alte abordări, cum ar fi transactivitatea. În [32] au fost analizate legăturile dialogistice.

Din punct de vedere al implementării, autorul a experimentat detectarea actelor de vorbire [46], a perechilor de adiacență, a lanțurilor lexicale și a celor dialogistice. Utilizarea însă a unor tehnici de învățare automată a fost făcută doar pentru detectarea actelor de vorbire (folosind modele Markov ascunse).

Pentru detectarea altor tipuri de legături implicite se va folosi sistemul TagHelper [47], dezvoltat la Universitatea Carnegie Mellon, care primește o conversație adnotată și care, după aplicarea unor pre-prelucrări tipice (tokenizare, stemming, POST etc.) încearcă să aplice mai mulți algoritmi de învățare automată pentru a obține un model al fenomenului studiat. Acesta va fi apoi folosit pentru adnotarea unor conversații neadnotate. Sisteme de învățare automată (WEKA - <http://www.cs.waikato.ac.nz/ml/weka/>, SVMLight - <http://svmlight.joachims.org/>)

Tehnicile de învățare automată folosite în sisteme care detectează legături implicite sunt:

1. perceptronul ("sequential Voted Perceptron Learner" al lui Collins),
2. Support Vector Machines (SVM),
3. Bayes naiv [48],

4. arbori de decizie (ID3, C4.5),
5. clasificare binară în cascadă (Confidence Restricted Cascaded Binary Classification)
6. învățare secvențială,
7. KNN,
8. modele Markov [54],
9. entropie maximă [49]
10. clustering [55]

În TagHelper [47], sistemul care se preconizează a fi folosit în cadrul temei, sunt folosite primele 6. Alte abordări de învățare automată în detectarea unor legături implicite sunt folosite în special în cazul coreferințelor (folosind entropia maximă [49], modele Markov [54]), a actelor de vorbire și a structurilor retorice (folosind Bayes naiv [48]).

Preprelucrări uzuale sunt:

- tokenizarea
- lematizarea sau extragerea rădăcinii (stemming)
- eliminarea cuvintelor nerelevante ("stopwords")
- găsirea cologațiilor
- adnotarea părților de vorbire (POST)
- găsirea entităților numite
- identificarea grupurilor nominale, verbale etc. ("chunking")

Trăsăturile considerate pentru învățare pot fi:

- semne de punctuație ("?", ",", ":", "!" etc.)
- lungimea intervenției (eventual normalizată)
- poziția intervenției în conversație
- distanța între replici
- proporția de numere în contribuție
- dacă sunt conținute cuvinte rare
- existența cel puțin a unui cuvânt nefuncțional [51]
- existența unui pronume
- existența la început a unor cuvinte cheie (da, nu, care, cine)
- este contribuția la subiect?
- reacțiile celorlalți la o rostire
- clasa (de exemplu, în WordNet) și rolul semantic
- apozitia (de exemplu, Băsescu, președintele României)
- este cel care intervine inițiatorul secvenței?
- este cel care intervine principalul contributor în secvență?
- coincidența între două intervenții
- paronimia
- similaritatea cosinus între două intervenții (personală și a celorlalți, eventual în grup)
- similaritate conform WordNet, Wikipedia sau a distanței Levenshtein
- distanța semantică în WordNet
- numărul sensului în WordNet

- familiaritatea din WordNet
- aliasuri (de exemplu, William și Bob, Nicolae și Nicu)
- importanța conform unui algoritm de tip PageRank
- informativitatea
- relevanța

După cum se vede, unele din aceste trăsături se referă strict la o intervenție iar altele consideră perechi adiacente [53] sau nu, sau chiar grupuri de intervenții. Pentru detectarea intervențiilor similare se poate folosi o măsură de tip cosinus, LSA, distanța Levensheim sau similarități între cuvinte via-WordNet, Wikipedia [49] sau alte astfel de resurse structurate.

Codificări ale legăturilor implicite în adnotare

O clasificare a dimensiunilor construirii cunoașterii în CSCL și codificarea asociată în analiza corpusurilor de interacțiuni este făcută de Weinberger și Fischer [50]:

- participare, aspecte:
 - cantitative – de exemplu, număr de cuvine
 - eterogenitate (participare in/egală) - agregarea deviației standard a valorilor cantitative a participării raportată la nivelul grupului
- epistemică – dacă și cum se lucrează la subiect și dacă se aplică adecvat conceptele
 - construirea spațiului problemei – legături între datele problemei pentru înțelegerea acesteia
 - construirea spațiului conceptual – legături între conceptele teoretice pentru înțelegerea teoriei
 - construirea legăturilor între spațiul problemei și spațiul conceptual
- argumentare
 - la nivel micro - pe baza teoriei lui Toulmin [40], în care se identifică: afirmația (*claim*), baza cu justificarea (*ground, warrant*), și calificarea (*qualifier*)
 - simplă
 - calificată – cu limitare a validității
 - cu bază ("grounded") – cu furnizarea unei baze
 - calificată și cu bază
 - neargumentative – întrebări, mutări de coordonare, metadescrieri
 - la nivel macro
 - argument
 - contra-argument
 - integrare
 - ne-argumentative
- moduri sociale – grade de trans-activitate [52]
 - externalizare – către membrii grupului
 - elicitare – utilizarea partenerului ca o resursă, de obicei prin punerea de întrebări [51]
 - consens rapid – acceptare pentru continuare
 - consens integrativ
 - consens către conflict
 - convergență, divergență

În [47] se propun 7 dimensiuni de analiză, care le includ pe cele anterioare:

- epistemică (35 categorii)
- micro-argumentare (4)
- macro-argumentare (6)
- sociale (21)
- reacție la elicitare și construire consensuală (3)
- răspunsuri la ”prompts” în mediul de învățare (4)
- contribuții noi sau citări (2)

Altă clasificare este cea a transactelor, făcută de Berkowitz și Gibbs:

	Mode = Non-competitive	Mode = Competitive
Primary Focus = Ego	* Feedback Request * Clarification	* Competitive Clarification * Refinement
Primary Focus = Alter	* Paraphrase * Justification Request * Completion * Extension	* Competitive Paraphrase * Contradiction * Reasoning Critique * Competitive Extension * Counter Consideration
Primary Focus = Dyad	* Juxtaposition * Common Ground / Integration * Dyad Paraphrase	* Competitive Juxtaposition * Comparative Critique

De exemplu, un transact ”Competitive Juxtaposition” [52] este:

A: well ...u do know increasing tmax and pmax mean more Qin

B: yeah - but more quality - which means you get more workout of the turbine

În analiza conversațiilor există și alte abordări de identificare a legăturilor între replici, cum ar fi cea a lui Hmelo-Silver [60] sau Law și ceilalți [59].

Categorie	Marcheri de discurs	Exemple
Reason (rațiune)	because, since	<i>Since</i> we rely on the supply from the mainland...
Condition (condiție)	if	<i>If</i> we don't have enough land,....
Consequence (consecință)	then, thus, so, therefore	<i>So</i> why don't we find some places where has more land...
Contrast (diferență)	but, though, although, however, even, otherwise	<i>Although</i> some of the pollutants blocked out some sunlight...
Elaboration (elaborare)	moreover, such as	<i>Moreover</i> , we can use reusable energy such as wind power
Claim (afirmație)	I think, I agree, We should	<i>I think</i> that HK Government should set up laws and make a random inspection
Question (întrebare)	What, Why, How	<i>Why</i> can spaceflight help humans develop a better world?
Rebuttal (dezacord)	"I don't think", "I don't agree", "we shouldn't"	<i>I don't think</i> budgets are serious problem.

Cohesion (coeziune)	Also, besides	CO2 can be produced with chemical method, <i>also</i> , we may have new technology...
---------------------	---------------	---

Tabelul 1: Categoriile de acte de vorbire argumentative și cuvintele cheie (marcheri de discurs - "discourse markers") utilizate pentru identificarea lor

Hmelo-Silver [60] introduce diagramele CORDTRA, (Chronologically-oriented Representations of Discourse and Tool-related Activity), în care identifică următoarele tipuri de replici:

Tabelul 2, categoriile CORDTRA

	Categorie	Exemple
Conținut	Task talk (discuție asupra sarcinii)	I recommend that you spend a bit more time on discussing EACH proposal and then vote let's say late afternoon on Tuesday.
	Tool-related (discuție asupra instrumentelor)	Frank and I decided instant messenger may be useful for discussing comprised info and ideas.
	Concept talk (discuție asupra conceptelor)	Elaborative rehearsal better equips the student with the information he is rehearsing because it becomes more accessible in his long term memory he has found ways to relate it to other instances and in his own words and he can help his peers understand it on a more simple level.
	Personal talk (discuție personală)	Hey, I just wanted to let everyone know that I will might be a little late logging in on Monday morning. I will be in Connecticut until early Monday morning.
Colaborare	New Ideas (idei noi)	Peer assessment done by each student in each group on their group members.
	Modifications (modificări)	I don't think its necessary to peer evaluate within the groups. We might try to give roles out within the group to make sure that each student has a part in the experiment and is working and not slacking off.
	Agreement (accord)	I like Mary's proposal for a hypothesis sheet...
	Disagreement (dezacord)	I don't believe peer assessment should be a factor in the student's grade but it could be done as feedback for both students and teachers to use.
	Summaries (rezumate)	What we have so far: Jack – teacher beliefs, Beth – hands on learning, Ellen - prior knowledge use, Carol – cognitive flexibility theory, Sylvia – collaborative learning
	Acknowledgment (înștiințare)	Helen did this too so I just want to emphasis a few main concepts and points.
Întrebări	Informational	Should we meet before class at 9:15 so we can go over and refine what we have done?
	Explanation	What do you mean by self-regulated learning?
	Metacognitive	What do others of you think?
Complexitate	Telling (povestire)	Direct instruction: method of instruction for mastery of basic skills, concepts, strategies, facts, and information. This instruction is done piece by piece rather than all together. ...
	Elaborated telling (povestire elaborată)	Games and activities that students are familiar with can help teach specific facts about a country its culture, and its language. ... I thought Monopoly would be a good game to use. As long as it was carefully coded as to appropriate linguistic level and maturity level suggested for students.
	Transforming (transformare)	This idea supports our objective of "Transfer knowledge of static electricity to everyday examples" And so both motivation and transfer can be achieved through Authentic Instruction, two characteristics of which are Students' work has value beyond the school setting. Lessons become more authentic as the connection to the real world is increased. ... Thus, one (authentic) activity would be a field experience (trip) to a lab, power plant, etc., which would should real world use and is social.
Justificări	Personal beliefs (credeințe personale)	I think a good assessment is to have each student do a mini science project based on static electricity, which I think would encourage students to think more and to not just concentrate on a grade.
	Grounded beliefs	According to Sociocultural theory, "to capture a student's motivation, the culture of

	Categorie	Exemple
	(credințe împământenite)	school must find a way to be valuable, relevant, interesting, and challenging in the eyes of a child. This may mean engaging the students in authentic activities of the larger society. It also means challenging them with tasks that are meaningful to the larger culture and are relevant to their lives outside of the school environment." ...
Monitorizare	Individual monitoring (individuală)	I have made a summary about the stuff I got from the knowledge web that i posted as well as my research and printed it out so we can attach it to our sticky paper on Thursday.
	Group monitoring (de grup)	Ok, so I think we need to revise or come to a concensus (sic) about how we want to word our final proposals.
	Self-directed learning (învățare auto-dirijată)	I will research metacognition. I still need to look up the concept of "self-directed learning"
	Other monitoring (altele)	Let's meet on Monday after class to talk about our gallery walk.

O posibilitate de adnotare a actelor de vorbire este limbajul DAMSL (Dialog Act Markup in Several Layers [62]), care codifică pe diferite nivele informațiile referitoare la replicile unui dialog. Două dintre aceste nivele, *funcția de anticipare* (FLF, în engleză, "forward looking function") și *funcția de adaptare regresivă* (BLF în engleză, "backward looking function") sunt extensii ale actelor de vorbire ce extrag noțiuni legate de structura dialogului cum ar fi perechile de adiacență.

Funcția de anticipare leagă replica curentă de o replică următoare pe baza actelor de vorbire de genul Searle/Austin [57,62]:

Afirmații	o afirmație făcută de cel ce vorbește
Cerere de informație	întrebare pusă de cel ce vorbește
Confirmare	întrebare pentru confirmarea unor informații
Influența asupra ascultătorului	= directivile lui Searl
Opțiuni	o sugestie slabă sau o listă de opțiuni
Ordin	o comandă
Influența asupra vorbitorului	= directivile lui Austin
Ofertă	vorbitorul se oferă să facă ceva și așteaptă confirmarea
Execuție	vorbitorul face ceva
Acte Convenționale	
Deschidere	mesaje de întâmpinare
Incheiere	mesaje de închidere
Mulțumire	multumiri și răspunsuri la mulțumiri

Funcția de adaptare regresivă se bazează pe relația dintre replica curentă și alte replici, care au fost enunțate anterior de către alte persoane. Aceasta include acceptarea/refuzarea propunerilor precum și alte structurări ale dialogului [57,62]:

Acord	răspunsul vorbitorului la o propunere anterioară
Acceptă	acceptarea propunerii
Acceptă Parțial	acceptă o parte a propunerii
Poate	nici nu acceptă, dar nici nu refuză
Refuză	refuzarea propunerii
Refuză Parțial	refuzarea parțială a propunerii
Abținere	amană/evită răspunsul
Răspuns	răspunsul la o întrebare
Înțelegere	dacă vorbitorul înțelege noțiunea anterioară
Semnal de neînțelegere	vorbitorul nu a înțeles
Semnal de înțelegere	vorbitorul a înțeles
Aprobare	demonstrare prin continuare

Repetare/Reformulare demonstrare prin repetiție/reformulare
Completare demonstrare prin completare prin colaborare

Strijbos&Xhafa :

Step 1	Step 2	Step 3	Step 4	Step 5
C-thread	Conversation	Social	<i>PS-thread</i>	Problem Solving
Reply to U_i	No code <i>State</i> Offer Request Regulate Repair typing Respond, <i>more general than the codes below that are tied to problem solving:</i> Follow <i>Elaborate</i> <i>Extend</i> <i>Setup</i> Agree Disagree Critique Explain	Identity self Identity other Interest Risk-taking Resource Norms Home School Collaborate group Collaborate individual Sustain climate Greet	<i>Connect to U_i</i>	Orientation Strategy <i>Tactic</i> Perform <i>Result</i> Check <i>Corroborate/counter</i> <i>Clarify</i> Reflect <i>Restate</i> Summarize

CONSTRUIREA CORPUSULUI ADNOTAT DE CONVERSAȚII

Predarea, dublată de implementarea și experimentarea tehnologiilor colaborative au fost avute în atenție de mai mulți ani la cursul de "Interfețe evolute", predat la anul terminal 5 (respectiv, din acest an, 4, fiind redenumit "Interacțiune om-calculator"), la specializarea Calculatoare a Facultății de Automatică și Calculatoare din Universitatea "Politehnica" București. Cursul acoperă capitolele clasice de interacțiune om-calculator (modelarea și evaluarea interfețelor, utilizabilitate, ergonomie cognitivă etc.), dar acordă un spațiu considerabil problematicii dezvoltării de sisteme interactive și a interfețelor specifice webului semantic, precum și a interfețelor inteligente și a celor destinate aplicațiilor colaborative, specifice Web2.0 (webul social). În cadrul orelor de aplicații se dau mai multe teme de casă, accentul fiind pe implementarea de interfețe avansate pentru web.

Utilizarea tehnicilor colaborative sunt folosite, de mai mulți ani, nu numai în cadrul cursului de Interacțiune om-calculator, ci și a cursurilor de Analiza algoritmilor (anul 3, respectiv 2), Complemente de informatică (anul 5) și Prelucrarea limbajului natural (master). În paralel cu activitățile didactice bazate pe colaborare au fost făcute și cercetări asupra modalităților în care se poate îmbunătăți procesul de învățământ folosind tehnologiile informatice colaborative. În acest sens au fost date teme de casă specifice acestor tehnologii și au fost proiectate, implementate și evaluate instrumente specifice de analiză semi-automată a temelor

de casă. Cercetările au făcut și obiectul a mai multor teme de cercetare la Institutul de Cercetări în Inteligența Artificială al Academiei Române și a unor granturi naționale, din care amintim Sintec [5, 6, 26] și K-Teams [28, 32], precum și internaționale (LarFLaST [1, 21-23, 35], COOPER [14,15], VMT [32] și LTfLL [16, 32]).

Din 1994 până în 2003 a fost investigată paradigma sistemelor inteligente de instruire (în engleză, "Intelligent Tutoring Systems", ITS), bazate pe prelucrarea cunoștințelor și ontologii [11, 19-21, 23, 26, 35]. Au fost făcute analize ale activității studenților folosind pre-teste, post-teste, teste de dependență de context, de personalitate, de profil cognitiv și emoțional [11,12] sau analiza parcurgerii paginilor de web [13]. S-a încercat găsirea unor factori care pot fi folosiți pentru generarea personalizată de suport de curs [12, 20, 21, 23, 25, 26, 35] sau de trasee de instruire, care folosesc materiale de curs deja existente [6, 27]. Nu a fost generalizată folosirea acestor tehnici inteligente de instruire datorită faptului că, chiar și pentru un profesor specializat în inteligență artificială este dificil să construiască bazele de cunoștințe formale și a ontologiilor necesare generării amintite mai sus. În plus, multe din cunoștințele necesare, în special la discipline de la specialitatea calculatoare necesită nu numai cunoștințe declarative, care pot fi exprimate în ontologii și baze de cunoștințe, ci și cunoștințe procedurale (de exemplu, referitoare la dezvoltarea de programe) sau, din altă perspectivă, tacite, intuitive. Un alt rol important în înțelegere îl au și elementele imagistice, narațiunea [17], metaforele care, deși pot fi parțial tratate folosind tehnici cognitive, de exemplu, [22], sunt mai apropiate de o abordare socio-culturală [34], bazată pe etnometodologie [9] sau pe analiza imaginarului [33].

Ca urmare a considerentelor anterioare, din 2005 s-a trecut de la paradigma cognitivă, ce fundamentează ITS la paradigma socio-culturală ce fundamentează CSCL. Prin urmare, la cursul de Interfațare om-calculator s-a trecut la predarea unor teorii cum ar fi teoria activității [8], dialogismul [3] iar la laborator au fost introduse teme de casă specifice CSCL, bazate pe chat și bloguri.

O constantă a utilizării CSCL în ultimii ani la cursurile amintite mai sus de la Facultatea de Automatică și Calculatoare a fost colaborarea folosind mesageria instantanee (chat). Ca urmare a experiențelor din cadrul proiectului Virtual Math Teams (VMT [32]), începând din anul 2005 una din temele de casă a constat în una sau mai multe sesiuni chat nemoderate, (la VMT erau discuții moderate) de minim o oră, cu 3-8 membri, în care studenții aveau de dezbătut un subiect legat de curs. Ei au decis data și ora efectuării sesiunii (bineînțeles, având totuși un termen limită de efectuare a ei). În 2005-2006 au folosit ce mediu de chat au dorit, iar din anul următor au folosit sistemul ConcertChat, care oferă posibilitatea referirii explicite a replicilor anterioare sau a unor zone de pe spațiul grafic de lucru ("whiteboard") [10].

Înainte de prima sesiune de chat, în câteva cazuri, studenții au avut de căutat pe Web și de citit documentație referitoare la temele indicate. În majoritatea cazurilor au avut de făcut și câteva activități suplimentare după sesiunile de chat, care s-au încadrat în trei categorii: scrierea unor mici texte (rezumate sau reflecții), răspunsul la câteva întrebări sau adnotarea unor sesiuni de chat proprii (în 2007-2008) sau ale altor colegi (2008-2009). La sfârșitul fiecărei sesiuni de chat, studenții au trebuit să trimită pentru fiecare echipă înregistrarea ("logul") sesiunii chat. Tema de casă colaborativă pe chat a fost în toți anii opțională, dar a fost punctată cel puțin cu o zecime din nota finală, astfel încât participarea a fost mare (vezi

tabelul 1). Într-un caz au fost date și la examen 1-2 subiecte legate de chestiunile discutate în sesiunile de chat.

Din punct de vedere al cercetării, au fost investigate mai multe aspecte, cum ar fi:

- Construirea cunoașterii în echipe virtuale pe chat
- Atragerea studenților în activități colaborative folosind chat
- Analiza modului în care ei transmit cunoașterea de la o sesiune de chat la alta folosind narațiunea și implicit, în colaborarea ulterioară
- Rezumarea textelor (într-o temă de casă dată la cursul de Intefefe om-calculator) și a discuțiilor chat, de către studenți și automat
- Reflecția
- Transmiterea cunoștințelor de la o sesiune de chat la alta
- Argumentarea și negocierea în discuțiile chat
- Analiza interacțiunii și a secvențelor în care are loc o colaborare efectivă
- Identificarea legăturilor implicite între replici
- Proiectarea și implementarea de sisteme de:
- Rezumare a chaturilor
- Identificare a legăturilor implicite
- Identificarea secvențelor de construire a cunoașterii
- Vizualizarea firelor de discuție
- Recomandare de documente plecând de la conținutul chaturilor
- Colaborarea studenților folosind chat

An	Subiect de discuții	Activități suplimentare	Nr.total de studenți înscriși la curs	Nr. sesiuni	Nr. studenți în echipe	Distribuirea studenților	Nr. total de echipe	Nr. de studenți participanți
2005 - 2006	Facilitățile utile pentru un mediu de lucru colaborativ	1. Rezumare 2. Reflecție 3. Răspuns la întrebări	235	3	4	Fixată de profesor	47/33/50	200
2006 - 2007	Facilitățile utile pentru un mediu de lucru colaborativ	Reflecție	232	2	4	Fixată de profesor	65	159
2007 - 2008	Dezbateri wiki, chat, forum, blog, urmată de integrare	Adnotare		1	4	Fixată de profesor	55	173
2008 - 2009	Dezbateri wiki, chat, forum, blog, urmată de integrare	Adnotare	324	2	8	La alegerea studenților	27/30	160

Tabelul 3. Sinteza conținutului și a participării la sesiunile de chat

În anii universitari 2005-2006 și 2006-2007, subiectul dat studenților pentru dezbateri în temele de casă a fost discutarea facilitatilor care ar fi utile pentru un mediu de lucru colaborativ pentru rezolvarea de probleme sau proiectare, folosind texte (fara voce sau video in timp real). Această temă de casă s-a desfășurat în anul universitar 2005-2006 în trei pași (în anii următori numărul variind între unul și doi), la interval de circa două săptămâni, în echipe de câte patru fixate de profesor. De la un pas la altul echipele s-au schimbat, păstrând doi membrii și aducând alți doi (din echipe anterioare diferite). La fiecare pas ei au avut de făcut una din următoarele trei activități (și neavând de făcut aceeași activitate de două ori, astfel încât fiecare să facă una din cele 3 activități):

1. rezumare
2. răspuns la întrebările:
 - Ce idei utile ați descoperit în timpul sesiunii de chat, care ar putea fi transmise pentru pasul următor?
 - Ce probleme nerezolvate ar trebui abordate în pasul următor?
 - Ce ar trebui să facă echipa altfel la pasul următor ca să lucreze mai bine împreună?
 - Cât v-a ajutat rezumatul sau răspunsurile la întrebări pentru a înțelege ce s-a discutat în pasul 1? (întrebare pentru pașii 2 și 3).
3. Reflecție la conținutul sesiunii (părerii despre sesiune)

În anii 2008 și 2009, tema a început cu un studiu, studenții trebuind să caute pe web și să înțeleagă informațiile referitoare la avantajele, dezavantajele, instrumentele de sprijinirea interfațării cu bloguri, conferințe chat, forumuri de discuție și wiki, precum și posibilitățile lor de utilizare în lucrul colaborativ. După aceea, au trebuit să facă o conferință chat în echipe de câte patru (în 2009 echipe de până la nouă membri), folosind sistemul ConcertChat. Sesiunea de chat a trebuit să aibă două părți:

1. O discuție în contradictoriu despre instrumentele de colaborare pe web: conferințe chat, blog, wiki, forumuri de discuții. Pentru aceasta, fiecare dintre cei patru participanți și-a ales unul dintre cele 4 instrumente (în 2009, fiecare instrument a fost ales de câte doi participanți, permițând astfel să existe și o colaboare în argumentare), considerând că este "patronul" unei firme care comercializează astfel de instrumente și discută cu ceilalți "patroni" susținându-și produsul și criticându-le pe celelalte.
2. În a doua parte a aceleași sesiuni chat, membrii echipei au discutat cum s-ar putea face o integrare a celor 4 instrumente de colaborare într-o aplicație.

După efectuarea sesiunii de chat, la un interval de circa două-trei săptămâni, studenții au lucrat individual pentru a analiza și adnota chatul lor (în 2008) respectiv dintr-o sesiune a unor colegi pe care nu-i cunoșteau (în 2009). Ideea principală a fost de a adnota legăturile implicite (care nu au fost indicate cu ConcertChat) între replici, de a identifica și adnota secvențele de colaborare intensă, conceptele (topicele) introduse (în afara celor de start: wiki, forum, chat, blog) și de a da note. În acest scop a fost implementat un program interactiv de adnotare pus la dispoziție studenților [7].

Schema de adnotare

În decursul anilor în care s-au făcut experimentele, a fost dezvoltată o schemă de adnotare în XML, atât a componentelor chaturilor cât și a adnotărilor referitoare la legături implicite, note și regiuni de colaborare.

Schema de adnotare a chaturilor este:

```
<?xml version="1.0" encoding="UTF-16LE"?>
  <!ELEMENT Corpus (#PCDATA|Dialog)*>
  <!ELEMENT Dialog (Participants,Topics,Body)>
  <!ATTLIST Dialog team CDATA #REQUIRED>
  <!ELEMENT Participants (Person)+>
  <!ELEMENT Person (#PCDATA)>
  <!ATTLIST Person nickname CDATA #REQUIRED>
  <!ELEMENT Topics (Item)*>
  <!ELEMENT Item (#PCDATA)>
  <!ELEMENT Body (Turn)+>
  <!ATTLIST Turn nickname CDATA #REQUIRED>
  <!ELEMENT Turn (Utterance)+>
  <!ELEMENT Utterance (#PCDATA)>
  <!ATTLIST Utterance
    genid CDATA #REQUIRED
    time CDATA #REQUIRED
    ref CDATA #REQUIRED>
```

Un exemplu de chat adnotat automat conform acestei scheme, plecând de la fișierul HTML generat de ConcertChat este:

```
<Dialog team="sesiuneal_concertChat.html">
  <Participants>
    <Person nickname="mihS"/>
    <Person nickname="cezar"/>
    <Person nickname="Radu"/>
    <Person nickname="monica b"/>
    <Person nickname="micky"/>
    <Person nickname="ank"/>
  </Participants>
  <Topics/>
  <Body>
    <Turn nickname="mihS">
      <Utterance genid="1" ref="-1" time="14/12/2008 04:47:17">joins the room</Utterance>
    </Turn>
    <Turn nickname="cezar">
      <Utterance genid="2" ref="-1" time="14/12/2008 04:48:17">joins the room</Utterance>
    </Turn>
    <Turn nickname="Radu">
      <Utterance genid="3" ref="-1" time="14/12/2008 04:48:29">joins the room</Utterance>
    </Turn>
    <Turn nickname="cezar">
      <Utterance genid="4" ref="-1" time="14/12/2008 04:48:33">Cezar, 352 C4</Utterance>
    </Turn>
    <Turn nickname="monica b">
      <Utterance genid="5" ref="-1" time="14/12/2008 04:48:37">joins the room</Utterance>
    </Turn>
    <Turn nickname="micky">
      <Utterance genid="6" ref="-1" time="14/12/2008 04:48:50">joins the room</Utterance>
    </Turn>
    <Turn nickname="micky">
      <Utterance genid="7" ref="-1" time="14/12/2008 04:49:20">Mihaela, 351C3</Utterance>
    </Turn>
    <Turn nickname="ank">
      <Utterance genid="8" ref="-1" time="14/12/2008 04:49:35">joins the room</Utterance>
    </Turn>
    <Turn nickname="Radu">
      <Utterance genid="9" ref="-1" time="14/12/2008 04:49:41">Hello friends my name is Radu352 c2</Utterance>
    </Turn>
    <Turn nickname="monica b">
      <Utterance genid="10" ref="-1" time="14/12/2008 04:49:42">Monica - 352C2</Utterance>
    </Turn>
    <Turn nickname="mihS">
      <Utterance genid="11" ref="-1" time="14/12/2008 04:49:44">hello, my name is Mihaela and i'm interested in developing a better software for conference chat</Utterance>
    </Turn>
    <Turn nickname="mihS">
      <Utterance genid="12" ref="-1" time="14/12/2008 04:50:21">for now i'm a student at A&amp;C at UPB, 352C4</Utterance>
```



```

</Turn>
<Turn nickname="mihS">
  <Utterance genid="13" ref="-1" time="14/12/2008 04:50:34">what about the rest of
you?</Utterance>
</Turn>
<Turn nickname="ank">
  <Utterance genid="14" ref="-1" time="14/12/2008 04:50:51">hello, my name is Gabriela,
352C4</Utterance>
</Turn>
<Turn nickname="monica b">
  <Utterance genid="15" ref="-1" time="14/12/2008 04:51:11">hello, I'll talk about
wiki</Utterance>
</Turn>
<Turn nickname="Radu">
  <Utterance genid="16" ref="11" time="14/12/2008 04:51:12">hello mihaela i am really
interested in your theories</Utterance>
</Turn>
<Turn nickname="Radu">
  <Utterance genid="17" ref="-1" time="14/12/2008 04:51:35">in fact i doubt that this is the
right approach</Utterance>
</Turn>
.....

```

O schemă mai completă de adnotare, care a fost adoptată într-o primă instanță [45] este:



Figure 2. Schemă de adnotare

Adnotarea făcută manual, folosind editorul care va fi prezentat mai jos se face conform schemei următoare:

```

<?xml version="1.0" encoding="UTF-16LE"??>
<!ELEMENT annotation (topics,grades,collab_regions,implicit_refs,comments)>
<!ATTLIST annotation
  file CDATA #REQUIRED
  annotator CDATA #REQUIRED
  annotation_date CDATA #REQUIRED>
<!ELEMENT Topics (item)+>
<!ELEMENT Item (#PCDATA)>
<!ELEMENT Grades ((general_grade)+,(instant_grade)+)>
<!ELEMENT General_grade EMPTY>
<!ATTLIST General_grade
  collaboration CDATA ""
  ontopic CDATA ""
  nickname CDATA ""
  value CDATA "">
<!ELEMENT Instant_grade EMPTY>

```

```

<!ATTLIST Instant_grade
    nickname CDATA #REQUIRED
    genid_ref CDATA #REQUIRED
    value CDATA #REQUIRED>
<!ELEMENT Collab_regions (collab_region)*>
<!ELEMENT Collab_region EMPTY>
<!ATTLIST Collab_region
    start_genid_ref CDATA #REQUIRED
    end_genid_ref CDATA #REQUIRED>
<!ELEMENT Implicit_ref (from_text,to_text,type,pattern)>
<!ATTLIST Implicit_ref
    from_genid_ref CDATA #REQUIRED
    to_genid_ref CDATA #REQUIRED>
<!ELEMENT Implicit_refs (implicit_ref)+>
<!ELEMENT From_text (#PCDATA)>
<!ELEMENT To_text (#PCDATA)>
<!ELEMENT Type (#PCDATA)>
<!ELEMENT Pattern (#PCDATA)>
<!ELEMENT General_comment EMPTY>
<!ELEMENT Comments (general_comment)>

```

Un exemplu de astfel de adnotare este dat mai jos:

```

<Annotation annotation_date="20/01/2009 04:52:47" annotator="Alexandru, 27, 342C3">
  <Topics>
    <Item>Search engine</Item>
    <Item>new software</Item>
  </Topics>
  <Grades>
    <General_grade collaboration="10"/>
    <General_grade ontopic="10"/>
    <General_grade nickname="mihS" value="10"/>
    <General_grade nickname="cezar" value="10"/>
    <General_grade nickname="Radu" value="10"/>
    <General_grade nickname="monica b" value="10"/>
    <General_grade nickname="micky" value="10"/>
    <General_grade nickname="ank" value="10"/>
    <Instant_grade genid_ref="61-&gt;80" nickname="mihS" value="10"/>
    <Instant_grade genid_ref="61-&gt;80" nickname="cezar" value="10"/>
    <Instant_grade genid_ref="61-&gt;80" nickname="micky" value="10"/>
    <Instant_grade genid_ref="61-&gt;80" nickname="ank" value="9"/>
    <Instant_grade genid_ref="201-&gt;220" nickname="mihS" value="10"/>
    <Instant_grade genid_ref="201-&gt;220" nickname="micky" value="10"/>
    <Instant_grade genid_ref="201-&gt;220" nickname="ank" value="10"/>
    <Instant_grade genid_ref="60" nickname="ank" value="10"/>
    <Instant_grade genid_ref="74" nickname="ank" value="5"/>
    <Instant_grade genid_ref="75" nickname="monica b" value="10"/>
    <Instant_grade genid_ref="141" nickname="ank" value="10"/>
  </Grades>
  <Collab_regions>
    <Collab_region end_genid_ref="103" start_genid_ref="96"/>
    <Collab_region end_genid_ref="218" start_genid_ref="199"/>
  </Collab_regions>
  <Implicit_refs>
    <Implicit_ref from_genid_ref="11" to_genid_ref="12">
      <From_text>hello, my name is MIhaela and i'm interested in developing a better software
for conference chat</From_text>
      <To_text>for now i'm a student at A&amp;C at UPB, 352C4</To_text>
      <Type>elaboration</Type>
      <Pattern/>
    </Implicit_ref>
    <Implicit_ref from_genid_ref="12" to_genid_ref="13">
      <From_text>for now i'm a student at A&amp;C at UPB, 352C4</From_text>
      <To_text>what about the rest of you?</To_text>
      <Type>elaboration</Type>
      <Pattern/>
    </Implicit_ref>
    <Implicit_ref from_genid_ref="16" to_genid_ref="17">
      <From_text>hello mihaela i am really interested in your theories</From_text>
      <To_text>in fact</To_text>
      <Type>elaboration</Type>
      <Pattern>.. in fact</Pattern>
    </Implicit_ref>

```

```

<Implicit_ref from_genid_ref="35" to_genid_ref="36">
  <From_text>thie real time can be a real pain sometimes</From_text>
  <To_text>yes, but consider you wark at a project</To_text>
  <Type>elaboration</Type>
  <Pattern/>
</Implicit_ref>
<Implicit_ref from_genid_ref="44" to_genid_ref="45">
  <From_text>well the open style of a wiki can be useful for group projects, while a blog
may aid in brainstorming or generating discussion</From_text>
  <To_text>although i have to admit that a combination of the two may best suit
educational purposes.</To_text>
  <Type>contrast</Type>
  <Pattern>.. although</Pattern>
</Implicit_ref>
<Implicit_ref from_genid_ref="48" to_genid_ref="49">
  <From_text>what</From_text>
  <To_text>because</To_text>
  <Type>answer</Type>
  <Pattern>what .. ? - .. because</Pattern>
</Implicit_ref>
<Comments>
  <General_comment/>
</Comments>
</Annotation>

```

Editor pentru adnotarea chaturilor

Pentru adnotarea chaturilor, a fost dezvoltat un editor specializat [7], folosit deja de doi ani în cadrul cursurilor mai sus menționate pentru obținerea unui corpus adnotat. Aplicatia este realizata in Java 1.6 si integrează elemente grafice de tip SWING.

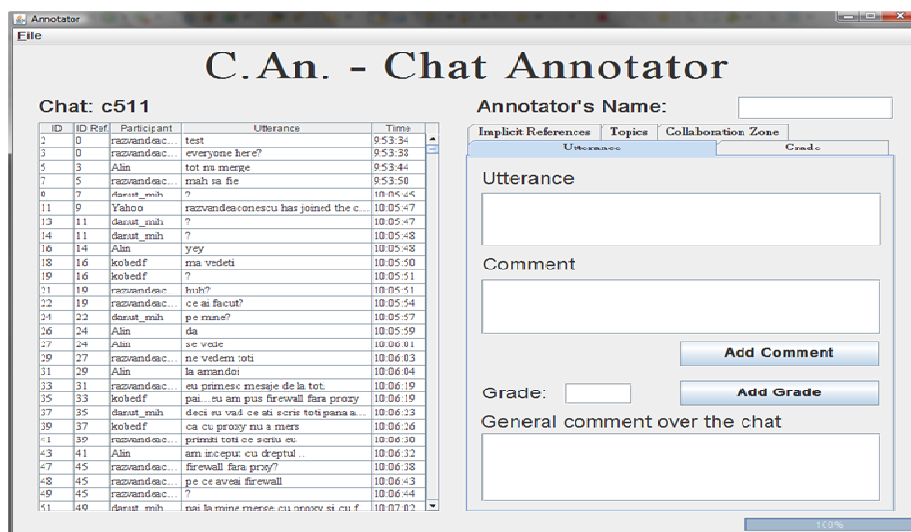


Figura 3. Interfața Adnotator

Aplicatia permite adnotarea mai multor informații cu privire la chat-ul analizat. În afară de identificarea de legături implicite se pot adnota și următoarele informații:

- comentarii, note pentru o replică oarecare;
- note pentru fiecare participant pentru o înșiruire de 20 de replici succesive, reflectând activitatea sa din respectiva zonă;
- note finale pentru participanții la chat și pentru relevanța chat-ului: colaborarea dintre participanți și rămânerea "on topic";
- în funcție de evoluția chat-ului, cuvintele cheie/relevante de pe parcurs (topic-urile);

- specificarea zonelor de colaborare intensă în care informațiile oferite sunt relevante raportat la discuția purtată

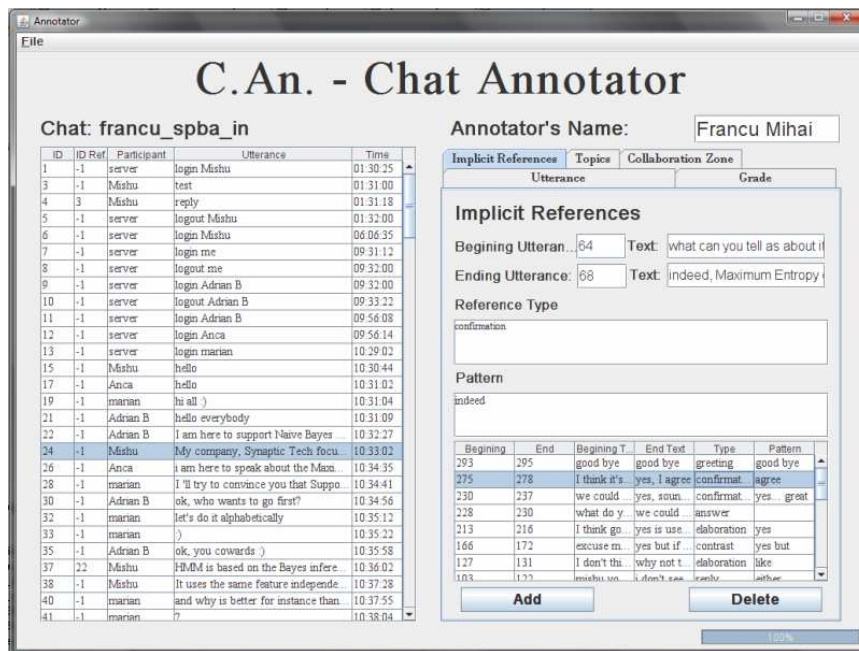


Figura 4. Exemplu de legături implicite

CONCLUZII ȘI DIRECȚII VIITOARE DE CERCETARE

Există multe abordări în învățarea automată a diverselor tipuri de legături implicite. În lucrare au fost trecute în revistă mai multe seturi de adnotări folosite precum și trăsături folosite în diverse aplicații.

A fost descris corpusul de conversații ”Polyphony”, schemele de adnotare folosite și, în detaliu, contextul în care a fost construit. De asemenea, a fost implementat un editor pentru adnotarea legăturilor implicite, experimentat timp de doi ani cu un număr mare de studenți.

În viitor este necesară identificarea unui set de adnotări care să fie folosite (din cele prezentate), precum și a unui set de trăsături. De asemenea, se vor face experimente cu TagHelper și alte sisteme.

O dimensiune importantă de investigat este definirea eventuală a unei modalități de adnotare a firelor de discuție și încercarea de a implementa ideile polifoniei bahntiniene, cu încercarea folosirii de tehnici de învățare automată și în această direcție.

REFERINȚE

1. G. Angelova, S. Boytcheva, O. Kalaydjiev, St. Trausan-Matu, P. Nakov, A. Strupchanska, Adaptivity in a web-based CALL system, in F. van Harmelen (ed.): Proceedings of ECAI-2002, the 15th European Conference on AI, IOS Press, Amsterdam, 2002, pp. 445-449
2. Marcu, D., The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts, PhD Thesis, Department of Computer Science, University of Toronto, Toronto, Canada, December 1997
3. M.M., Bakhtin, The Dialogic Imagination: Four Essays, University of Texas Press, 1981.
4. K. Bollacker, R. Cook, P. Tufts, Freebase: A Shared Database of Structured General Human Knowledge. Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada. AAAI Press 2007, p. 1962-1966
5. V. Cristea, St Trausan-Matu, O. Udrea, A Semantic Web Approach to e-learning Personalization, WSEAS Transactions On Information Science and Applications, Issue 5, Volume 1, November 2004, pp. 1320-1325.
6. V. Cristea, St Trausan-Matu, Authoring and delivering adaptable learning objects in SINTEC, 3rd Workshop of Authoring of Adaptive and Adaptable Educational Hypermedia, AIED Conference, Amsterdam, 2005 (vezi și http://www.wis.win.tue.nl/~acristea/AAAEH05/papers/8-vcristea_formatted.pdf, descărcată în mai 2009)
7. M. Dascalu, E. Chioasca, St. Trausan-Matu, ASAP- An Advanced System for Assessing Chat Participants, in D. Dochev, M. Pistore, and P. Traverso (Eds.): AIMSA 2008, LNAI 5253, Springer, 2008, pp. 58-68.
8. Y. Engeström, Innovative learning in work teams: Analysing cycles of knowledge creation in practice. În Y. Engeström, R. Miettinen, and R.L. Punamäki (Eds.), Perspectives on activity theory, Cambridge, UK: University Press, 1999, pp. 377-404.
9. H. Garfinkel, H. Studies in ethnomethodology. Englewood Cliffs, NJ: Prentice-Hall, 1967.
10. T. Holmer, A. Kienle, M. Wessner, Explicit Referencing in Learning Chats: Needs and Acceptance, în Innovative Approaches for Learning and Knowledge Sharing, First European Conference on Technology Enhanced Learning, Nejd, W., Tochtermann, K. (eds.), Lecture Notes in Computer Science, 4227, Springer, 2006, pp. 170-184
11. Gh. Iosif, I. Juvina, St. Trausan-Matu, A. Marhan, Aspecte ale achizitiei de cunostinte in modelarea studentului pentru un sistem inteligent de instruire, in Revista de psihologie a Academiei Romane, Bucuresti, Romania, Tomul 45, 1-2, 1999, pp.31-50
12. Gh. Iosif, St. Trausan-Matu, I. Juvina, A. Marhan, Factori psihologici si educationali implicati in instruirea asistata de sisteme inteligente, in Revista de psihologie a Academiei Romane, Bucuresti, Romania, Tomul 46, nr.1-2, 2000, pp. 55-64.
13. Juvina, Gh. Iosif, St. Trausan-Matu, G. van der Veer, A. Marhan, C. Chiselita, Analysis of Web Browsing Behavior -- a great potential for psychological research, Proceedings of TAMODIA 2002, pp. 170-177
14. V. Posea, D. Mihaila, St Trausan-Matu, V. Cristea, A. Gartner, Evaluation of Virtual Learning Environments Using Logs and Social Networks, 1st International Workshop on Building Technology Enhanced Learning solutions for Communities of Practice (TEL-CoPs'06), Crete, Greece, Sept. 2006.
15. V. Posea, St. Trausan-Matu, V. Cristea, Online Evaluation of Collaborative Learning Platforms, Proceedings of 1st International Workshop on Collaborative Open Environments for Project-Centered Learning, ECTEL2007, Crete, Greece, 2007, pp. 66-75.
16. T. Rebedea, St. Trausan-Matu, C. Chiru, Extraction of Socio-semantic Data from Chat Conversations in Collaborative Learning Communities, in P. Dillenbourg and M. Specht (Eds.): EC-TEL 2008, LNCS 5192, Springer, 2008, pp. 366-377.
17. J. Sarmiento, St Trausan-Matu, G. Stahl, Co-constructed narratives in online, collaborative mathematics problem solving. Proceedings of Narrative Learning Environments Workshop, 12th International Conference on AI in Education (AIED 2005), Amsterdam, Netherlands, pp. 56-62, 2005
18. G. Stahl, T. Koschmann, D. Suthers, Învățarea colaborativă sprijinită de calculator: o perspectivă istorică, in St. Trausan-Matu (ed.), Interacțiunea conversatională in sistemele colaborative pe web, Editura MatrixRom, 2008, pp. 1-24.
19. St. Trausan-Matu, Programe inteligente pentru asistarea invatarii, in Revista Romana de Informatica si Automatica, Bucuresti, Romania, vol.5, nr.4, 1995, pp. 7-16
20. St. Trausan-Matu, Knowledge-Based, Automatic Generation of Educational Web Pages, in Proceedings of Internet as a Vehicle for Teaching Workshop, Ilieni, iun. 1997, pp.141-148.

21. St. Trausan-Matu, Web Page Generation Facilitating Conceptualization and Immersion for Learning Finance Terminology, in N. Nistor și M. Jalobeanu (eds.), RILW99, 1999, pp. 112-122.
22. St. Trausan-Matu, Metaphor Processing for Learning Terminology on the Web, in Lecture Notes in Artificial Intelligence, Nr. 1904, Springer, 2000, pp. 232-241.
23. St. Trausan-Matu, Intelligent personalizing web pages and understanding facilities, in S.A.Cerri, D.Maraschi (eds.), Proceedings of WITREC-2000, Montpellier, France, pp. 59-68.
24. St. Trausan-Matu, Gh. Iosif, I. Juvina, A. Marhan, Task Modeling in Intelligent Tutoring Systems on the Web, Proceedings of TAMODIA 2002, pp. 18-24.
25. St. Trausan-Matu, A. Marhan, Gh. Iosif, I. Juvina, Generation of Cognitive Ergonomic Dynamic Hypertext for E-Learning, in Harris, D., Duffy, V., Smith, M., Stephanidis, C. Human-Centred Computing. Cognitive, Social and Ergonomic Aspects, Mahwah, NJ, USA, Lawrence Erlbaum Associates Publishers, Vol 3, 2003, pp. 607-611.
26. St. Trausan-Matu, V. Cristea, O. Udrea, Sisteme inteligente de instruire, Editura PolitehnicaPress, 2005.
27. St. Trausan-Matu, D. Mihaila, V. Cristea, Personalizarea secvențierii într-un mediu de învățare bazat pe cunoștințe, in C.G. Apostol, D. Gorgan (eds.), Interacțiune om-calculator 2006, Informatica Economica, vol. X, București, 2006, pp. 49-54.
28. St. Trausan-Matu, G. Stahl, J. Sarmiento, Polyphonic Support for Collaborative Learning, in Y.A. Dimitriadis et al. (Eds.): CRIWG 2006, Lecture Notes in Computer Science 4154, Springer, 2006, pp. 132 – 139
29. St. Trausan-Matu, Traian Rebedea, Alexandru Dragan, Catalin Alexandru, Visualisation of Learners' Contributions in Chat Conversations, in Fong., J., Wang, P. (Eds.), Blended Learning, Pearson Prentice Hall, 2007, pp. 215-224.
30. St. Trausan-Matu, Gerry Stahl, Johann Sarmiento, Supporting Polyphonic Collaborative Learning, E-service Journal, vol. 6, nr. 1, Indiana University Press, 2007, pp. 58-74.
31. St. Trausan-Matu, Polifonia colaborării dialogale, in St. Trausan-Matu (ed.), Interacțiunea conversatională în sistemele colaborative pe web, Editura MatrixRom, 2008, pp. 41-57.
32. St. Trausan-Matu, Traian Rebedea, Polyphonic Inter-Animation of Voices in VMT, in G. Stahl (ed.), Studying Virtual Math Teams, Springer, 2009. Vezi și <http://www.ischool.drexel.edu/faculty/gerry/vmt/book/24.pdf>, descărcată în mai 2009.
33. St. Trausan-Matu, Studiu asupra determinării legăturilor implicite în dialoguri, ICIA, iunie 2008
34. L. Trausan-Matu, V.A. Canciu, St. Trausan-Matu, Analiza imaginărilor din texte, in A. Marhan, I. Juvina (eds.), Proceedings la a 4-a Conferința Națională de Interacțiune Om-Calculator – RoCHI 2008, MATRIX ROM, București, pp.97-104
35. Vygotsky, L., Mind in society, Cambridge, MA: Harvard University Press, 1978
36. Trausan-Matu, St., Maraschi, D. and Cerri, St. (2002), Ontology-Centered Personalized Presentation of Knowledge Extracted From the Web, in S.Cerri, G.Gouarderes (eds.), Intelligent Tutoring Systems 2002, Springer, Lecture Notes in Computer Science number 2363, pp 259-269.
37. Suthers, D. D., Dwyer, N., Medina, R., & Vatrappu, R. (2006). A Framework for Analyzing Interactional Processes in Online Learning. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Chicago, April 9-13, 2007.
38. Mahesh Joshi and Carolyn Penstein Rosé. Using Transactivity in Conversation for Summarization of Educational Dialogue, www.cs.cmu.edu/~cprose/pubweb/SLaTE07-JoshiM.pdf, 12 iunie 2008
39. Tannen, D., Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse, Cambridge University Press, 1989
40. Toulmin, S., *The Uses of Arguments*. Cambridge Univ. Press, 1958
41. Sacks, H., Lectures on conversation. Oxford, UK: Blackwell, 1992.
42. Duranti, Al., Linguistic Anthropology, Cambridge University Press, 1997
43. Mann și Thompson, Rhetorical structure theory: Toward a functional theory of text organization, 1988
44. Stahl, G., Group Cognition: Computer Support for Building Collaborative Knowledge, MIT Press, 2006.
45. Ciprian Onofreiciuc, Alexandru Rosiu, Alexandru Gartner, Stefan Trausan-Matu, Polyphony, a Knowledge-based Chat System Supporting Collaborative Work, in C. Badica, M. Paprzycki (Eds.), Advances in Intelligent and

Distributed Computing, Proceedings of IDC 2007 - First International Symposium on Intelligent and Distributed Computing, Studies in Computational Intelligence Vol. 78, Springer, 2007, pp. 155-164

46. Chiru Trausan-Matu S., Chiru C. și Bogdan R. (2004). Identificarea actelor de vorbire în dialogurile purtate pe chat, în Trausan-Matu S., Pribeanu C. (Eds.), *Interacțiune Om-Calculator 2004*, Editura Printech, Bucuresti, pp. 206-214
47. Rose, C. P., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F. (In Press). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning, *International Journal of Computer Supported Collaborative Learning*
48. Teufel S., Moens, M., Summarizing scientific: Experiments with Relevance and Rhetorical Status, *Computational Linguistics*, 28 (4), 2002
49. Ponzeto, S.P., Strobe, M., Exploiting Semantic Role Labeling, WordNet and Wikipedia for Coreference Resolution, Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL, pages 192–199, New York, June 2006.
50. Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46, 71-95.
51. Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, 17 (4), 416-426
52. Joshi, M., Rose, C.P., Using Transactivity in Conversation for Summarization of Educational Dialogue <http://www.cs.cmu.edu/~cprose/pubweb/SLaTE07-JoshiM.pdf>, June 2009
53. Purandare, A., Litman, D., Analyzing Dialog Coherence using Transition Patterns in Lexical and Semantic Features, <http://www.amruta-purandare.com/pubs/flairs08-1.pdf>, June 2009
54. Poon, H., Domingos, P., Joint Unsupervised Coreference Resolution with Markov Logic, <http://www.cs.washington.edu/homes/pedrod/papers/emnlp08.pdf>, June 2009
55. Ng, V., Cardie, C., Improving Machine Learning Approaches to Coreference Resolution, <http://acl.ldc.upenn.edu/P/P02/P02-1014.pdf>, June 2009
56. Hobbs, J.R., On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University, 1985.
57. Juravsky, D., Martin, J., *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Prentice Hall: San Francisco 2000
58. Manning, C., Schütze, H., *Foundations of Statistical Natural Language Processing*, MIT Press: Cambridge (Mass.) 1999
59. Law, N., Lu, J., Leng, J., Yuen, J., și Lai, M., Understanding Knowledge Building from Multiple Perspectives, în Proceedings of Workshop on Interaction Analysis at ICLS, Utrecht, 2008.
60. Hmelo-Silver, E.C., Description of Chronologically-oriented Representation of Discourse and Tool Related Activity, în Proceedings of Workshop on Interaction Analysis at ICLS, Utrecht, 2008.
61. Boboc, Al., *Limbaj și ontologie*, Ed. Didactică și Pedagogică, 1997
62. Allen, J., Core, M., Draft of DAMSL: Dialog Act Markup in Several Layers, <ftp://ftp.cs.rochester.edu/pub/packages/dialog-annotation/manual.ps.gz>
63. Hirst, M., G. & Budanitsky, A. (2001). Lexical Chains and Semantic Distance. Proceeding Euroalan-2001, Iași.