ReaderBench goes Online: A Comprehension-Centered Framework for Educational Purposes

Gabriel Gutu, Mihai Dascalu, Stefan Trausan-Matu
University Politehnica of Bucharest
313 Splaiul Independentei, Bucharest, Romania
gabriel.gutu@cs.pub.ro, mihai.dascalu@cs.pub.ro, stefan.trausan@cs.pub.ro

Philippe Dessus
Univ. Grenoble Alpes, LSE
F-38000 Grenoble, France
philippe.dessus@univ-grenoble-alpes.fr

ABSTRACT
In this paper we introduce the online version of our *ReaderBench* framework, which includes multi-lingual comprehension-centered web services designed to address a wide range of individual and collaborative learning scenarios, as follows. First, students can be engaged in reading a course material, then eliciting their understanding of it; the *reading strategies* component provides an in-depth perspective of comprehension processes. Second, students can write an essay or a summary; the *automated essay grading* component provides them access to more than 200 textual complexity indices covering lexical, syntax, semantics and discourse structure measurements. Third, students can start discussing in a chat or a forum; the *Computer Supported Collaborative Learning* (CSCL) component provides in-depth conversation analysis in terms of evaluating each member's involvement in the CSCL environments. Eventually, the *sentiment analysis*, as well as the *semantic models and topic mining* components enable a clearer perspective in terms of learner's points of view and of underlying interests.

Author Keywords
Sentiment analysis; Semantic models; Topic mining; Automated essay grading; Reading strategies; Computer Supported Collaborative Learning; *ReaderBench* framework.

ACM Classification Keywords
I.2.7 [Natural Language Processing]: Discourse, Language parsing and understanding, Text analysis.

General Terms
Algorithms, Design, Languages.

INTRODUCTION
Knowledge understanding from texts, either read or written, are crucial in education-centered contexts. Technology has gained a broader usage and more tools designed to support tutors and learners alike in the learning process are being made available nowadays. Thus, a huge amount of content is being generated by teachers who share their learning materials, or by students who provide feedback, do tests, homework or are involved in online conversation.

Natural Language Processing (NLP) techniques [1] have gained considerable ground lately as they provide accurate and efficient analyses of both written and oral language. Advanced NLP services are being developed, including the analysis of unstructured learning materials of students' textual traces, automated essay grading, sentiment analysis, concept map elaboration or identification of reading strategies. Our framework, *ReaderBench* [2, 3, 4, 5], comprises of advanced NLP techniques used to expose a wide variety of language services. We can consider our framework as being unique as it provides a unitary core engine centered on cohesion and on dialogism [6, 7], the latter being reflected in the implemented polyphonic model [8]. Multiple connected services addressing different facets of comprehension assessment and prediction are thus deployed. Tutors are capable to perform an apriori assessment of learning materials, but also to evaluate a posteriori learner's written traces consisting of essays, self-explanations or utterances in CSCL conversations. All these services are described in detail in subsequent sections.

A client-site web application for our framework was being developed within the H2020 RAGE (Realising and Applied Gaming Eco-System) project, covering most back-end *ReaderBench* functionalities, and is currently available online at http://readerbench.com. Figure 1 depicts the main interface of the website.

This paper presents an overview of the online version of our framework regarding the services currently made available. Enhanced functionalities are still under development, while some web services were specifically implemented to meet RAGE partner requirements. A full web version that enables a holistic analysis of texts in general and of CSCL conversation, similar to the desktop application, will be made available in the foreseeable future.

In terms of structure, the second section introduces the overall *ReaderBench* architecture, while the third section presents in detail all language services that are currently published online. The fourth section presents specific use cases, as well as conclusions and future work.

ARCHITECTURE
The *ReaderBench* framework integrates a wide variety of advanced NLP techniques centered on comprehension assessment and prediction and is built around Cohesion Network Analysis [9]. *ReaderBench* has introduced a multi-lingual and automated model applicable to various types of texts, such as essays, self-explanations or conversations in Computer Supported Collaborative Learning (CSCL) environments and represents a framework that aims to reach targeted education purposes. Therefore, a variety of linguistic features

important for understanding texts and predicting learners' comprehension are made available. These include sentiment analysis, textual cohesion and textual complexity. In terms of inputs, besides plain text, some services use PDF files from which the extracted raw text is sent for processing. Other types of inputs, such as *Word* documents or RTF files will be considered in the nearest future.

As an overview, the *ReaderBench* framework makes use of the Standard Core NLP [10] for implementing natural language processing pipelines consisting of the following processes [1]: tokenization, sentence splitting, part of speech tagging, lemmatization, named entities recognition, dependency parsing, and co-reference resolution. Whereas for English the full pipeline is supported, for other languages (e.g., French, Spanish, Italian, Romanian and Dutch) only the core steps are being performed. In addition, *ReaderBench* includes multiple libraries such as Apache Mahout (http://mahout.apache.org/), Gephi (http://gephi. org/), and Mallet (http://mallet.cs.umass.edu/).

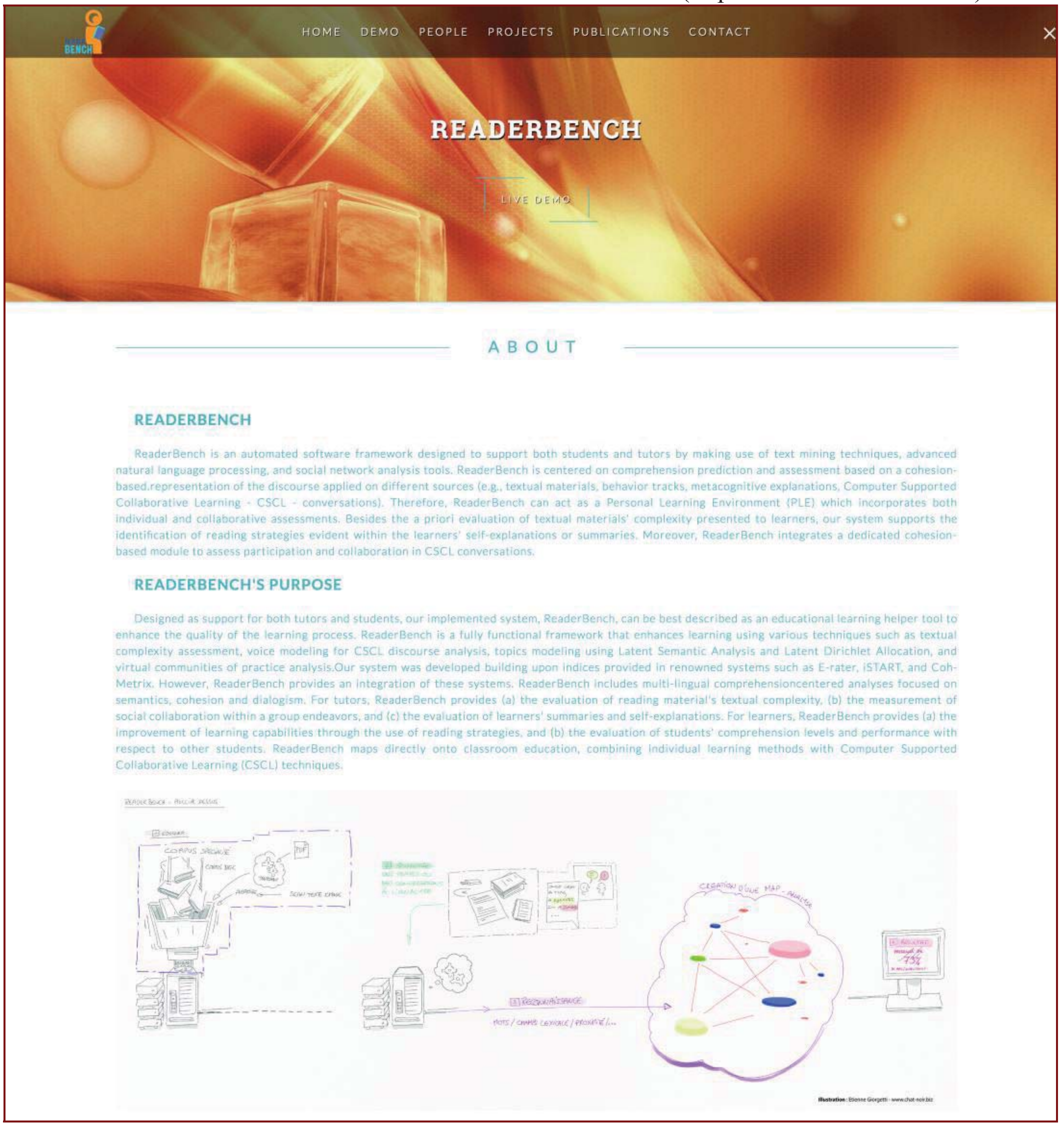

Figure 1. *ReaderBench* main web interface.

Cohesion is evaluated from multiple perspectives within the framework [11] in terms of semantic distances in lexicalized ontologies (e.g., WordNet, WOLF for French) [12], Latent Semantic Analysis (LSA) [13], and Latent Dirichlet Allocation (LDA) [14] semantic models. The models were trained on specific text corpora. Some of the corpora used for English language include Touchstone Applied Science Associates, Inc. corpus (TASA) (http://lsa.colorado.edu/spaces.html), the LAK dataset [15], or the Contemporary American English collection (COCA) [16]. Some of the texts used for French language include the Texts Enfants collection [17] and "Le Monde" corpus (http://lsa.colorado.edu/spaces.html). Figure 2 depicts the five most important components being included within the framework. The underlying services will be further described in the next sections.

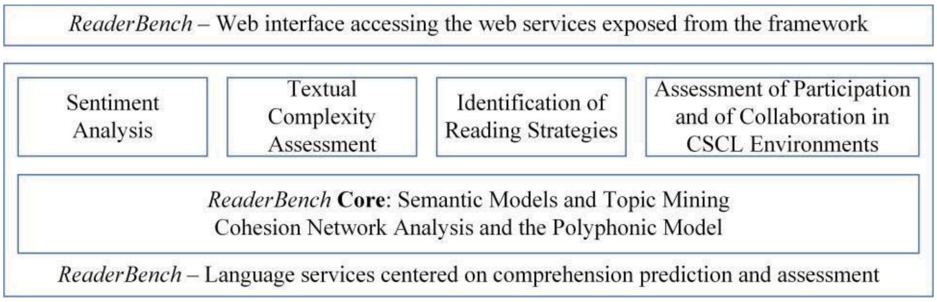

Figure 2. *ReaderBench* architecture.

LANGUAGE SERVICES

Five major components are currently presented within the web interface of the *ReaderBench* framework. Each component presented below is currently available as REST web services, and can be integrated in custom applications. JSON format is used for both sending data and accepting responses for the majority of our web services.

Automated Identification of Reading Strategies

Identification of reading strategies is a recognized predictor in determining the reading comprehension of students [18]. This component is also available on the *ReaderBench* website and it can be used to automatically identify metacognition, causality, bridging, paraphrasing and elaboration strategies used by a learner within their self-explanation [19].

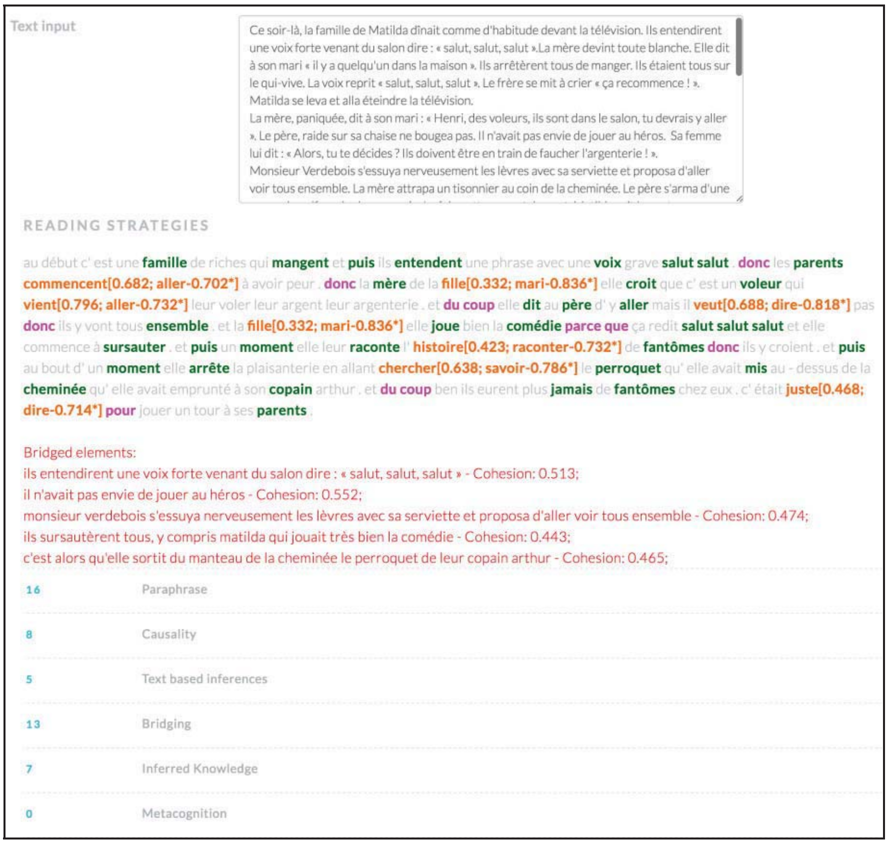

Figure 3. Sample input data for *ReaderBench* self-explanation service and automatically identified reading strategies.

Further analyses consider the usage of textual complexity indices in order to improve the accuracy in terms of comprehension prediction [20]. Figure 3 depicts a different sample input for French language. Based on a given target text, learners self-explain what they understood and specific employed reading strategies are automatically identified.

Textual Complexity Assessment

Automated essay grading represents a technique used to reduce tutor's workload by offering specific analyses and statistics regarding students' writing style. The model for textual complexity assessment, centered on cohesion and integrated in the *ReaderBench* framework, represents the foundation for a multi-dimensional analysis on writing styles. The generated indices support tutors in identifying improvements that can be done on each student's essay and enable an objective evaluation of students by offering them automatically generated feedback, which has a positive impact on writing style quality [21].

Some of the complexity indices reflected through our web service include statistic surface indicator (e.g., average paragraph, sentence or word lengths, number of commas, word and character entropy), syntax factors (statistics on different parts of speech, average number of first, second or third person pronouns per paragraph, depth of parsing tree), semantic cohesion (intra- and inter- paragraph and sentence cohesion scores computed using Wu-Palmer semantic distance over WordNet [22], LSA and LDA).

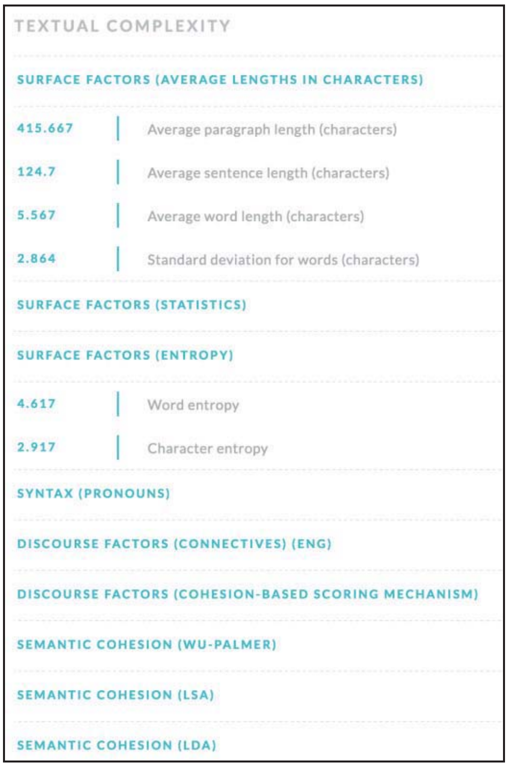

Figure 4. Textual complexity results computed for the sample input data.

Choosing an appropriate text for students, neither too simple nor too difficult to understand, represents an important task in the learning process. The indices provided by our tool are an important component when it comes to adapt learning materials for specific students. Valuable feedback can be retrieved by analyzing and combining the previous textual complexity indices all-together, thus supporting comprehension both a priori during text selection, as well as a posteriori during automated feedback generation. Figure 4 shows the textual complexity index scores obtained for the previous input data.

Automated Assessment of Participation and Collaboration in CSCL Conversations

Computer-Supported Collaborative Learning (CSCL) gains a broader usage due to technology adoption, while dialogism represents the most adequate framework for representing CSCL conversations [23, 24]. Concurrently, the need for automated conversation analysis tools to support tutors in the cumbersome process of analyzing students' interactions and activity has increased. Collaboration, which can be viewed as the inter-animation of ideas or opinions pertaining to different participants, represents a central element of dialogue [8]. Several analyses performed based on our CSCL collaboration evaluation models [9] are available on the website. These include participant interaction scores with an interaction graph built on top of Cohesion Network Analysis and visually displayed using the D3.js library. Specific indices are being computed for each participant, such as: number of contributions, cumulated contribution scores, degree of inter-animation, cumulative social knowledge building scores, in- and out- degree, closeness, betweenness, and eccentricity centrality measures from the interaction graph, relevance for top 10 conversation topics [9].

Each participation and collaboration index is used for obtaining an in-depth perspective of each member's involvement, followed by specific visual graphs. The first graph from Figure 5 depicts each participant's evolution as cumulative contribution scores across the timeframe of the conversation. The following two graphs depict the collaboration between participants in terms of the social knowledge building and the voice inter-animation model. Spikes with these 2 graphs denote intense collaborations spanning throughout the conversation.

In terms of underlying computational processes, the importance of each contribution is first computed by relying on the relevance of the covered topics from the entire conversation and present within the utterance. Second, collaboration was computed as the impact on other members' contributions in terms of cohesion (a longitudinal analysis of the conversation) and dialogism (a transversal analysis based on co-occurrence voice patterns). Therefore, within these models, collaboration was assessed using a bottom-up approach which emphasize that cohesion is a signature of collaboration.

Sentiment Analysis

Sentiment analysis and opinion mining are often referred in linguistic and psychological research in recent years. The sentiments extracted from author's text (for example, participants' contributions in a conversation or the absence of their interaction) provide information regarding author's feelings. Interaction established between members in a conversation influences further contributions and interactions.

The analysis of the participants' sentiments can take into consideration specific optimizations, such as ignoring contributions that do not cover specific topics or excluding contributions with no further references or irrelevant regarding main topics.

Specific goals can be defined given a text in terms of sentiment analysis. For example, specific sentiments from an input text can be extracted and split into the 6 major categories expressed by Picard [25]: excited, sad, scared, angry, tender and happy. A demo showing this approach is available on the *ReaderBench* website. In the backend, the framework computes these major sentiments combining scores for valences gathered from specific lists. English, French and Dutch languages are currently supported.

Figure 5. CSCL graphs generated for a sample conversation file.

Figure 6 shows an example of sentiment analysis results produced by the framework for the previous sample input. Negative results express absolute values for negative emotions, therefore emphasizing the positive nature of the entire text.

Figure 6. Sentiment analysis results computed for the sample input data.

The common resource for all considered languages represents the Linguistic Inquiry and Word Count (LIWC) dictionary [26] which contains words related to psychological phenomena, personal concerns, thoughts, feelings, personality, and motivations. At present, all dictionaries are used to explore their linguistic coverage and only those that are present in at least 20% of entry samples are considered for follow-up statistical analyses. The following word dictionary lists were integrated for English language in an approach similar to the one proposed by Crossley et al. [27]:

- Affective Norms for English Words (ANEW) [28], which provides values on three dimensions (valence, arousal and dominance) for more than 1,000 English verbs, nouns, and adjectives;
- Geneva Affect Label Coder (GALC) [29], which contains affective valences such as admiration, amusement, anger, anxiety and many others;
- EmoLex [30], comprising sentiments like anger, anticipation, disgust, fear and others;
- SenticNet [31], including five affective norms: pleasantness, attention, sensitivity, aptitude and polarity;
- Harvard IV-4 from the *General Inquirer* (GI) [32], which contains valences such as power, weak, active, passive, legal and more others;
- Lasswell dictionary [33], which includes sentiments like power gain, power loss, affective gain, affective loss and some others.

In addition, the Affective Norms for French words (FAN) [34] and the Dutch Affective Word Norms [35], the equivalent French and Dutch versions of ANEW, are also integrated in *ReaderBench*.

Semantic Models and Topic Mining

For this core component, the *ReaderBench* framework uses semantic similarity metrics based on ontologies (e.g., Wu-Palmer distance applied on WordNet), as well as cosine similarity between LSA word vectors and the inverse of the Jensen-Shannon dissimilarly between LDA topic distributions [11].

Cohesion Network Analysis introduced a generalized model based on the cohesion graph to represent discourse structure and underlying cohesive links. Based on CNA, a topic mining module was implemented, which extracts the most relevant concepts from a text. Integrated within the web interface, this module draws a concept map of these keywords: the nodes represent the central topics and the links between them depict the semantic similarity between two concepts; the size of each node is proportional to its relevance. Figure 7 presents the obtained concept map for a given input text, which is used for all subsequent print-screens for English language.

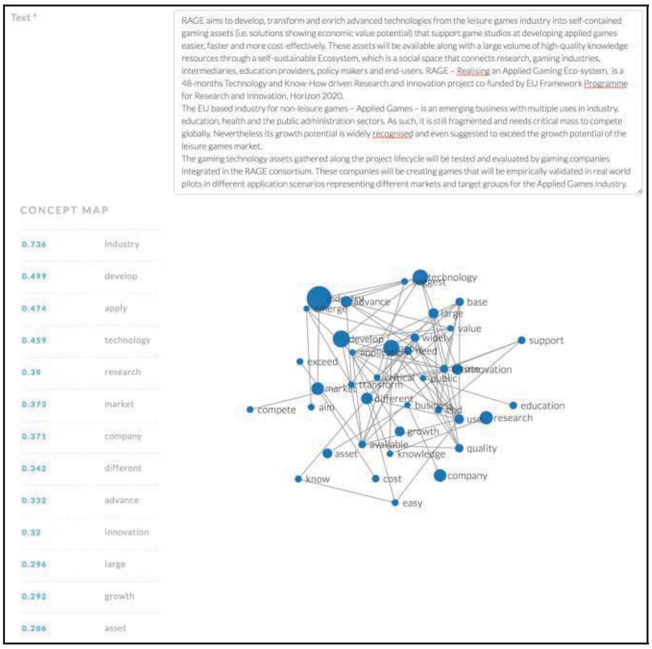

Figure 7. Sample input data for the *ReaderBench* web interface and the corresponding generated concept map.

EDUCATIONAL SCENARIOS

Up until recently, the desktop version of our *ReaderBench* framework was hardly usable in hands-on educational contexts due to the requirements of extensive processing power and high amounts of memory usage. Due to these limitations, it was mostly used in follow-up offline analyses. The online version opens up new usages of *ReaderBench* in education, as our framework can now be effectively used in a wide range of educational situations and needs. First, students can be engaged in reading a course material, then eliciting their understanding of it. *ReaderBench* can *identify their reading strategies*, providing an in-depth perspective of comprehension processes used to obtain a coherent mental representation of discourse.

Second, students can write an essay or a summary integrating the content of diverse topics from the course material. The *automated essay grading* component provides them access to more than 200 textual complexity indices integrated within a multi-layered model that covers lexical, syntax, semantics and discourse structure measurements.

Third, students can start discussing the course topics in a CSCL environment (chat, forum or blog). The *Computer Supported Collaborative Learning* (CSCL) component is centered on conversation analysis in terms of automated indices of participation and of collaboration, essential for evaluating each member's active involvement in the discussion. Eventually, the *sentiment analysis* component detects positive and negative emotions expressed in texts that, corroborated with the *semantic models and topic mining* component, enable a clearer perspective in terms of points of view and of underlying interests.

Besides this overall scenario, specific educational experiments were undergone in order to validate our models. Some of them are available online on our *ReaderBench* website, while others were built only for specific analyses and were not published online as web services. For example, of particular interest, is a serious game, currently under development, that enables users to enter textual competitions (e.g., creativity mini-games to identify inferred concepts, essay writing contests, self-explanations covering specific reading strategies) with other learners and to win based on higher predicted comprehension scores. Advanced techniques may be used to group students into clusters and the teaching material could be differentiated for each group. Another particular example of an extension currently under development is a tool focused on a contextual CV analysis. Given a PDF file representing a personal CV, the tool extracts specific indices and applies specific statistic model in order to predict whether the CV is adequate or not.

As future functionality enhancements, besides the Principal Component Analysis used to identify representative dimensions for each corpus in terms of sentiment analysis, specific improvements are also considered: integration of rules for valence shifting and the consideration of only positive and negative reviews, disregarding neutral or irrelevant content.

As a concluding remark, we must emphasize the extensibility of our *ReaderBench* framework and its broad potential usage in terms of integration within education scenarios performed in various languages. This paper is specifically meant to provide a global overview of the developed web interface, whereas specific details and validations are presented in detail in referred papers.

ACKNOWLEDGMENTS

This work was funded by the 644187 EC H2020 RAGE (Realising and Applied Gaming Eco-System) http://www.rageproject.eu/project.

REFERENCES

1. Manning, C.D. and Schütze, H. *Foundations of* statistical *Natural Language Processing*. MIT Press, Cambridge, MA, 1999.
2. Dascalu, M., Stavarache, L.L., Dessus, P., Trausan-Matu, S., McNamara, D.S., and Bianco, M. ReaderBench: The Learning Companion. In *17th Int. Conf. on Artificial Intelligence in Education (AIED 2015)* Springer, Madrid, Spain, (2015), 915–916.
3. Dascalu, M., Stavarache, L.L., Dessus, P., Trausan-Matu, S., McNamara, D.S., and Bianco, M. ReaderBench: An Integrated Cohesion-Centered Framework. In *10th European Conf. on Technology Enhanced Learning*, G. Conole, T. Klobucar, C. Rensing, J. Konert and É. Lavoué Eds. Springer, Toledo, Spain, (2015), 505–508.
4. Dascalu, M., Stavarache, L.L., Trausan-Matu, S., Dessus, P., Bianco, M., and McNamara, D.S. ReaderBench: An Integrated Tool Supporting both Individual and Collaborative Learning. In *5th Int. Learning Analytics & Knowledge Conf. (LAK'15)* ACM, Poughkeepsie, NY, (2015), 436-437.
5. Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., and Nardy, A. Mining texts, learner productions and strategies with ReaderBench. In *Educational Data Mining: Applications and Trends*, A. Peña-Ayala Ed. Springer, Cham, Switzerland, (2014), 335–377.
6. Dascalu, M., Trausan-Matu, S., Dessus, P., and McNamara, D.S. Discourse cohesion: A signature of collaboration. In *5th Int. Learning Analytics & Knowledge Conf. (LAK'15)* ACM, Poughkeepsie, NY, (2015), 350–354.
7. Dascalu, M., Trausan-Matu, S., Dessus, P., and McNamara, D.S. Dialogism: A Framework for CSCL and a Signature of Collaboration. In *11th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 2015)*, O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine and S. Ludvigsen Eds. ISLS, Gothenburg, Sweden, (2015), 86–93.
8. Trausan-Matu, S., Stahl, G., and Sarmiento, J. Supporting polyphonic collaborative learning. *Indiana University Press, E-service Journal 6*, (1), (2007), 58–74.
9. Dascalu, M., Trausan-Matu, S., McNamara, D.S., and Dessus, P. ReaderBench – Automated Evaluation of Collaboration based on Cohesion and Dialogism. *International Journal of Computer-Supported Collaborative Learning 10*, (4), (2015), 395–423.
10. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., and McClosky, D. The Stanford CoreNLP Natural Language Processing Toolkit. In *52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* ACL, Baltimore, MA, (2014), 55–60.
11. Dascalu, M. *Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence*. Springer, Cham, Switzerland, 2014.
12. Budanitsky, A. and Hirst, G. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics 32*, (1), (2006), 13–47.
13. Landauer, T.K. and Dumais, S.T. A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review 104*, (2), (1997), 211–240.
14. Blei, D.M., Ng, A.Y., and Jordan, M.I. Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, (4-5), (2003), 993–1022.
15. Taibi, D. and Dietze, S. Fostering Analytics on Learning Analytics Research: the LAK Dataset. In CEUR WS Proceedings Vol. 974, Proceedings of the LAK Data Challenge, held at LAK2013 – 3rd International Conference on Learning Analytics and Knowledg, Leuven, BE, 2013.
16. Davies, M. The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English. *Literary and Linguistic Computing 25*, (4), (2010), 447-465.
17. Denhière, G., Lemaire, B., Bellissens, C., and Jhean-Larose, S. A semantic space for modeling children's semantic memory. In *Handbook of Latent Semantic Analysis*, T.K. Landauer, D.S. McNamara, S. Dennis and W. Kintsch Eds. Erlbaum, Mahwah, (2007), 143-165.
18. McNamara, D.S., Levinstein, I., and Boonthum, C. iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers 36*, (2), (2004), 222–233.
19. Dascalu, M., Dessus, P., Bianco, M., and Trausan-Matu, S. Are Automatically Identified Reading Strategies Reliable Predictors of Comprehension? In *12th Int. Conf. on Intelligent Tutoring Systems (ITS 2014)*, S. Trausan-Matu, K.E. Boyer, M. Crosby and K. Panourgia Eds. Springer, Honolulu, USA, (2014), 456-465.
20. Dascalu, M., Stavarache, L.L., Dessus, P., Trausan-Matu, S., McNamara, D.S., and Bianco, M. Predicting Comprehension from Students' Summaries. In *17th Int. Conf. on Artificial Intelligence in Education (AIED 2015)* Springer, Madrid, Spain, (2015), 95-104.
21. Crossley, S.A., Allen, L.K., and McNamara, D.S., in press. The Writing Pal: A writing strategy tutor. In *Handbook on Educational Technologies for Literacy*, S.A. Crossley and D.S. McNamara Eds. Taylor & Francis, Routledge, New York, NY.
22. Wu, Z. and Palmer, M. Verb semantics and lexical selection. In *32nd Annual Meeting of the Association for Computational Linguistics, ACL '94* ACL, New Mexico, USA, (1994), 133–138.
23. Koschmann, T. Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In *Int. Conf. on Computer Support for Collaborative Learning (CSCL'99)*, C.M.

Hoadley and J. Roschelle Eds. ISLS, Palo Alto, (1999), 308-313.

24. Stahl, G., *Studying Virtual Math Teams*. Springer, New York, NY, 2009.

25. Picard, R. *Affective computing*. M.I.T Media Laboratory Perceptual Computing, 1997.

26. Pennebaker, J.W., Booth, R.J., and Francis, M.E. Linguistic inquiry and word count: LIWC [Computer software]. In *Austin, TX: liwc. net* University of Texas, Austin, TX, 2007.

27. Crossley, S., Kyle, K., and McNamara, D.S., (in press) Sentiment Analysis and Social Cognition Engine (SEANCE): An Automatic Tool for Sentiment, Social Cognition, and Social Order Analysis. *Behavior Research Methods*.

28. Bradley, M.M. and Lang, P.J. *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings.* The Center for Research in Psychophysiology, University of Florida, 1999.

29. Scherer, K.R. What are emotions? And how can they be measured? *Social science information 44*, (4), (2005), 695-729.

30. Mohammad, S.M. and Turney, P.D. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence 29*, (3), (2013), 436–465.

31. Cambria, E., Grassi, M., Poria, S., and Hussain, A. Sentic computing for social media analysis, representation, and retrieval. In *Social Media Retrieval*, N. Ramzan, R. Zwol, J.S. Lee, K. Clüver and X.S. Hua Eds. Springer, New York, NY, (2013), 191–215.

32. Stone, P., Dunphy, D.C., Smith, M.S., Ogilvie, D.M., and associates, *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press, Cambridge, MA, 1966.

33. Lasswell, H.D. and Namenwirth, J.Z., *The Lasswell Value Dictionary*. Yale University Press, New Haven, 1969.

34. Monnier, C. and Syssau, A. Affective norms for french words (FAN). *Behavior Research Methods 46*, (4), (2014), 1128-1137.

35. Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A.-L., De Schryver, M., De Winne, J., and Brysbaert, M. Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods 45*, (1), (2013), 169-177.