Revista Romana de Interactiune Om-Calculator 11 (1) 2018, 76-91

Liftoff – *ReaderBench* introduces new online functionalities

Gabriel Gutu-Robu[1], Maria-Dorinela Sirbu[1], Ionut Paraschiv[1], Mihai Dascalu[1,2], Philippe Dessus[2], Stefan Trausan-Matu[1,2]

[1] University Politehnica of Bucharest
313 Splaiul Independentei, Bucharest, Romania
E-mail: gabriel.gutu@cs.pub.ro, maria.sirbu@cti.pub.ro, ionut.paraschiv@cs.pub.ro, mihai.dascalu@cs.pub.ro, stefan.trausan@cs.pub.ro

2 Cognos Business Consulting S.R.L.

35 Blv Marasesti. sector 4, 040251, Bucharest, Romania

[3] Univ. Grenoble Alpes, LaRAC
F-38000 Grenoble, France
E-mail: philippe.dessus@univ-grenoble- alpes.fr

Abstract. Natural Language Processing (NLP) became a trending domain within recent years for many researches and companies due to its wide applicability and the new advances in technology. The aim of this paper is to introduce an updated version or our open-source NLP framework, *ReaderBench* (http://readerbench.com/), designed to support both students and tutors in multiple learning scenarios that encompass one or more of the following dimensions: Cohesion Network Analysis of discourse, textual complexity assessment, keywords' extraction using Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA) and word2vec semantic models, as well as the analysis of online communities and discussions. The latest version of our *ReaderBench* framework (v4.1) includes: a) new features, Application Programing Interfaces (APIs) and visualizations (e.g., sociograms, analysis of interaction between participants inside a community), b) a new web interface written in Angular 6, and c) the integration of new technologies to increase performance (i.e., spaCy and AKKA), as well as modularity and ease of deployment (i.e., Artifactory and Maven modules). *ReaderBench* is a fully functional framework capable to enhance the quality of learning processes conducted in multiple languages (English, French, Romanian, Dutch, Spanish, and Italian), and covering both individual and collaborative assessments.

Keywords: ReaderBench Framework, Natural Language Processing, Semantic Models, Cohesion Network Analysis, Computer-Supported Collaborative Learning.

1. Introduction

Natural Language Processing (NLP) gained an increasing popularity within the last years, both for researchers and the private sector. Researchers focus their efforts into developing algorithms and discovering new means to improve the accuracy of mechanisms designed to understand natural language. Companies feel the need to integrate services capable to analyze natural language from multiple data sources such as: online reviews, e-mails, conversations discussed within conference platforms or instant messaging applications. An example could consist of the automated analysis of online meetings or conversations between employees, to discover their topics of discussion, the expressed sentiments, or their degree of involvement throughout the conducted tasks. Companies may also need to gather data with regards to the perceived market impressions; thus, the analysis of customers' feedback is beneficial for taking timely decisions. Such scenarios are implemented with the help of NLP techniques, which are shown to provide efficient analyses of written texts (Manning & Schütze, 1999). Examples of practical applications in the NLP field cover also: automated machine translation, question answering systems, information extraction and interpretation services, as well as sentiment analysis services.

ReaderBench is an advanced NLP framework that integrates a variety of multi-lingual services aimed to analyze textual content in order to extract valuable information (Dascalu, 2014; Dascalu, Dessus, Bianco, Trausan-Matu, & Nardy, 2014). The framework relies on Cohesion Network Analysis (CNA) (Dascalu, McNamara, Trausan-Matu, & Allen, 2018) that evaluates text cohesiveness based on semantic similarity between units of text, for different granularities. CNA is a core constituent of our framework that supports follow-up analyses which may include: extraction of topics from texts, computing textual complexity indices indicative of specific writing traits and useful for performing automated essay scoring, analysis of different conversations (e.g., chat or forums) and processing of online communities.

The *ReaderBench* website incorporates demos for supported services, which are exposed through Representational State Transfer (REST) Application Programing Interfaces (APIs). The website was recently migrated from AngularJS to Angular 6, which is one of the latest versions developed by Google. The aim of this paper is to introduce the updated version (v4.1) of our framework with new features, such as: visualizations of online communities (including weekly sociograms), new integrated technologies (spaCy – https://spacy.io/ for text processing, Artifactory –

https://jfrog.com/artifactory/ for artefact management, Maven modules – https://maven.apache.org/, and AKKA – https://akka.io/ for concurrent processing), all targeting increased performance and scalability. *ReaderBench* exposes multiple APIs used by *RederBench* website and other systems, such as *ReadMe* (Botarleanu, Dascalu, Sirbu, Crossley, & Trausan-Matu, 2018; Sirbu, Botarleanu, Dascalu, Crossley, & Trausan-Matu, 2018) or third party developers (e.g., BipMedia for Randstad France within the Job Quest Game implemented in the H2020 RAGE project – Realising an Applied Gaming Ecosystem - http://rageproject.eu).

The structure of this article is as follows. The next section presents other existing NLP frameworks and applications, followed by details about the *ReaderBench* framework, integrated services and corresponding demos. After discussing educational scenarios (both envisioned and already conducted), the conclusions and future work section encourages scientific researchers and developers to perform their own experiments or to develop their own applications that rely on the services exposed by the *ReaderBench* API.

2. State of the Art

Reading and writing are central elements of learning, which are usually key factors in educational environments. Artificial Intelligence and Machine Learning algorithms play an important role in automatizing part of these processes, which are usually manually performed by humans (e.g., tutors in an academic environment). Natural Language Processing has gained a lot of popularity and it incorporates parts of the previously mentioned two domains, but also specific mechanisms aimed to understand and conceptualize the semantics of a text.

Multiple frameworks and libraries that incorporate NLP techniques have been developed and maintained, for example GATE (Cunningham, Maynard, Bontcheva, & Tablan, 2002), the Stanford CoreNLP library (Manning et al., 2014), spaCy – https://spacy.io/, and SyntaxNet – https://opensource.google.com/projects/syntaxnet. *GATE* is an open-source framework capable to support several NLP tasks in multiple languages, such as parsers and tagging, morphology, Information Retrieval and Extraction tools. It is a mature framework (more than 15 years old) which has been employed in multiple projects requiring text analysis.

Stanford CoreNLP (https://stanfordnlp.github.io/CoreNLP/) is an NLP toolkit written in Java that can perform linguistic analysis on texts written in multiple languages: English, French, German and Spanish (Manning et al., 2014). CoreNLP is also available in other programming languages such as Python. Because every language has its limitation, the full stack of functionalities are available only for English language. CoreNLP uses annotators to allow developers to specify the required modules they need to use. Some of the annotators, which are also used in *ReaderBench* framework, include: tokenizer and the sentence splitter, part-of-speech tagger, lemmatizer, named entity recognizer, dependency parser, and sentiment analysis. Stanford CoreNLP was integrated in previous *ReaderBench* versions, but was changed with spaCy (later on described) due to its increased robustness and processing speed (https://spacy.io/usage/facts-figures#speed-comparison).

SyntaxNet (https://ai.googleblog.com/2016/05/announcing-syntaxnet-worlds-most.html) is an open-source syntactic parser that was developed by Google. It covers an essential step in Natural Language Understanding, which is a subtopic in NLP that deals with machine reading comprehension. SyntaxNet is incorporated within TensorFlow, an open-source machine learning and deep learning framework able to perform high performance computations. For English, SyntaxNet uses a pre-trained model called *Parsey McParseface*, for whom a 94% accuracy was reported, a little higher than CoreNLP's with 93% accuracy (Chen & Manning, 2014).

SpaCy (https://spacy.io/) is a Natural Language Processing toolkit written in Python that supports deep learning algorithms. According to their official website, independent research showed that spaCy is the fastest NLP parser currently available. It incorporates a very fast syntactic parser aimed for production usage. SpaCy also integrates pre-trained models for part-of-speech tagging and dependency parsing for multiple languages. The updated version of our NLP framework integrates spaCy for all pre-processing analyzes.

3. The *ReaderBench* Framework

ReaderBench is an advanced open-source NLP framework centered on different facets of comprehension assessment and prediction. *ReaderBench* services are focused on semantics and in-depth text analyses with the aim is to stimulate individual and collaborative learning through comprehension prediction and assessment of written texts. *ReaderBench* fully supports multiple languages – English, French, Spanish, Dutch, Romanian –, while some languages (e.g., Italian and Latin) are only partially supported in terms of pre-processing stages. A pre-processing pipeline implemented on top of spaCy is applied on input texts and consists of multiple steps, out of which we mention tokenization, splitting, lemmatization, POS tagging, dependency parsing, named entity recognition, and the co-reference resolution.

ReaderBench has as central constituent text cohesion, a main linguistic feature indicative of discourse structure. Cohesion can be derived from discourse connectors like cue words or phrases (e.g., "but", "because"), referencing expressions, as well as lexical and semantic similarity between concepts. Within *ReaderBench*, cohesion is determined as an average semantic similarity measure of relatedness between textual segments that can be words, phrases, contributions, or even the entire document.

Cohesion Network Analysis (CNA) (Dascalu, Trausan-Matu, McNamara, & Dessus, 2015) builds a cohesion graph to represent discourse structure as cohesive links within a text. *ReaderBench* relies on semantic models and lexicalized ontologies (i.e., WordNet(s) for different languages) (Fellbaum, 2005) to build the cohesion graph. The incorporated semantic models include Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997), Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003), and word embeddings through word2vec (Mikolov, Chen, Corrado, & Dean, 2013).

The following subsections describe in detail the services which are available online on the *ReaderBench* website.

4.1. Open Online Services

The *ReaderBench* website incorporates demos for part of the services integrated in the framework. The website was previously running on AngularJS (https://angularjs.org) (Gutu, Dascalu, Trausan-Matu, & Dessus, 2016), a web application framework based on JavaScript and developed by Google. The website was recently migrated to Angular 6 (https://angular.io) (also known as Angular 2+). Our new website integrates new demo services and provides updated versions of the old services.

For English experiments, semantic models are trained using the following two corpuses: Touchstone Applied Science Associates (TASA) - http://lsa.colorado.edu/spaces.html, and The Corpus of Contemporary American English (COCA) - https://corpus.byu.edu/coca/. TASA is a large collection of documents that contains general texts, thus covering a large majority of the common words of the English language. For French language, "Le Monde" corpus was used - http://catalog.elra.info/en-us/repository/browse/ELRA-W0015/, which contains a collection of news from a popular French newspaper.

4.2. Sentiment Analysis

Sentiment analysis or opinion mining has shown an increasing adoption in marketing; however, few experiments are focused on games reviews (both serious and educational). *ReaderBench* integrates a sentiment analysis component that is able to extract the polarity of game reviews which was particularly useful in the context of the H2020 RAGE project. One practical scenario for the service may be the assessment of the reviews provided by users of a gaming website to understand how they perceive the game. Another scenario could be the real-time analysis of chat conversations that take place in online games, to assess participants' sentiments with regards to the interaction with the game and with other users. Multiple word lists are integrated in *ReaderBench*, enabling users to observe and compare emerging patterns. For English language, the following word lists are incorporated:

- ANEW (*Affective Norms for English Words*) (Bradley & Lang, 1999) – contains three affective norms (valence, arousal and dominance).
- GALC (*Geneva Affect Label Coder*) (Scherer, 2005) – contains affective valences such as admiration, amusement, anger, anxiety, boredom, compassion and many others.
- GI (*General Inquirer*) (Stone, Dunphy, Smith, Ogilvie, & associates, 1966) – recurrent patterns within the rich variety of man's written and spoken communications.
- General opinion lexicons (Hu & Liu, 2004) – contains lists of negative and positive words.
- Lasswell (Lasswell & Namenwirth, 1969) – contains sentiments like power gain, power loss, affective gain, affective loss and some

others.

- NRC Emotion Lexicon (Mohammad & Turney, 2013) – list of English words and their associations with eight basic emotions and two sentiments (negative and positive).

For French, the FAN (*Affective Norms for French*) (Monnier & Syssau, 2014) word list is used, while for Dutch the *Dutch Affective Word Norms* (Moors et al., 2013) was integrated.

The sentiment analysis service extracts word valences from these predefined word lists and the positive/negative scores are hierarchically weighed based on the relevance at paragraph/document level, via Cohesion Network Analysis. With the help of the sentiment analysis tool, users' perceptions with regards to a product may be now automatically assessed. Game companies can now better understand user experiences in serious games from a linguistic point of view, and potentially summarize opinions towards a specific feature. Applications in the educational field may include the assessment of discussions that take place in a Massive Open Online Course (MOOC) platform, with the aim of understanding student feelings towards a specific topic. Thus, the Sentiment Analysis service has applicability both in industry and educational environments by enabling the assessment of reactions towards a product or a service.

4.3. Automated Essay Scoring

Essay Scoring is usually seen as a time-consuming task and most existing systems rely on shallow analyses. Our *ReaderBench* framework tries to overcome this limitation by integrating a wide range of textual complexity indices that facilitate custom analyses for specific language scenarios. Our multi-dimensional model includes multiple indices categorized into several categories that address different dimensions of writing: surface analysis and classic readability formulas as baseline, syntax, semantics, discourse structure, as well as word complexity spanning across multiple levels (Dascalu, Crossley, McNamara, Dessus, & Trausan-Matu, 2018). In contrast to other systems, *ReaderBench* is centered on cohesion, both local and global, derived from CNA.

Practical applications of Automated Essay Scoring (AES) system include the assessment of the difficulty of a given text, and scoring mechanism for essays, summarizations, questions and self-explanations computed using models derived from textual complexity indices. As a follow-up extension within the *ReadMe* system, comprehensive feedback is provided with the main goal of presenting texts of adequate/steadily increasing complexity levels to learners. Moreover, AES systems facilitate the delivery of adequate texts in terms of complexity by considering learners' levels. Students and teachers are the main target and this model allows also the development of standardized tests.

4.4. Keywords Extraction

A topic mining module was implemented in *ReaderBench* using the Cohesion Network Analysis. The module extracts the most relevant concepts of a given text, with corresponding relevance scores and semantic relations with other concepts. The extracted concepts are visualized using a concept map in which the keywords represent the nodes, while the links between nodes illustrate the semantic similarity between two keywords. The keywords extraction demo allows users to enter a free-input text and perform topic mining on it. The most relevant keywords of the text, together with a concept map are displayed in an interactive way using a graph representation (see Figure 13). The size of a node is proportional to the corresponding keyword's relevance. The length of a link is inversely proportional with the semantic relatedness score between the two linked nodes. Figure 13 shows a preview of the results for the RAGE project description in which words like "industry", "develop", "research", "technology" and "market" were extracted as being the most relevant. The relations with other words are also emphasized on the concept map, thus a strong relevance between "industry" and "technology" can be observed. Connection with words that are not so relevant can also be notice, for example the pair "research" – "knowledge".

Our service has multiple applications both in businesses and educational environments. Extracting the main topics of a text enables entities to understand the main subjects of a given document. In educational environments, the extraction of keywords can be used in a chat conversation platform and identify whether participants deviate from the main topic.

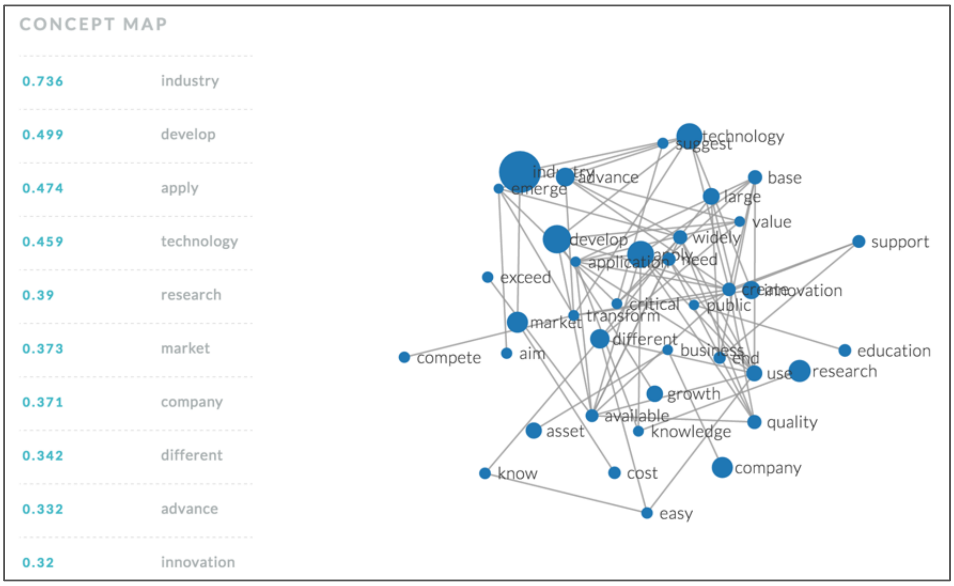

Figure 13. Keywords extraction results provided for the RAGE project description.

4.5. Curriculum Vitae Analysis

Curriculum Vitae (CV) Analysis is a service aimed to enhance the recruiting process by automatically scoring and categorizing CVs using multiple NLP metrics. This service may decrease recruitment time necessary to find suitable candidates by analyzing and exploring multiple CVs. Therefore, the selection of the most adequate candidate(s) can be simplified by automatically scoring CVs and selecting the ones which have the highest chance to fit a position. This service is designed as a human-aided decision system in which the final decision of choosing the proper candidates is, of course, performed by the recruiter.

The CV analysis service provides recommendations and feedback on two dimensions: textual content and visual aspect. The formulas behind the service were developed using a collection of French CVs provided by Randstad (Gutu et al., 2018), who annotated the collection into “positive” and “negative” CVs. Several features that included word lists and textual complexity indices were used to train classifiers that capture the specificities of the two dimensions. The tool currently provides feedback messages aimed at helping people who apply for a management position in order to improve the quality of their CV and increase their chances of acceptance.

4.6. Automated Model of Comprehension

The Automated Model of Comprehension (AMoC) service (Dascalu, Paraschiv, McNamara, & Trausan-Matu, 2018) models human reading by outlining the way in which a person discovers and conceives concepts, while reading a text. The service uses two data types: text-based data such as syntactic dependencies that link words from the input text, and inferred concepts, that are semantically related and are generated using the incorporated semantic models.

AMoC aims is to simulate comprehension. Understanding and simulating the reading process is a main feature for creating contextualized learning environments that stimulate the acquisition of new information. AMoC uses multiple NLP techniques to annotate unstructured text and highlight the most central concepts, both text-based and inferred, based on corresponding activation scores. Understanding and activating the most important words across multiple sentences can be used to assess the complexity of a text, and then provide readers with more detailed explanations. In educational environments, this service can also help students with their learning experiences by allowing them to easily understand unstructured information. Figure 14 shows the evolution of words, together with their activation scores, throughout a text that was used to demonstrate de Landscape Model (http://www.cogcrit.umn.edu/docs/vandenBroek_10.shtml).

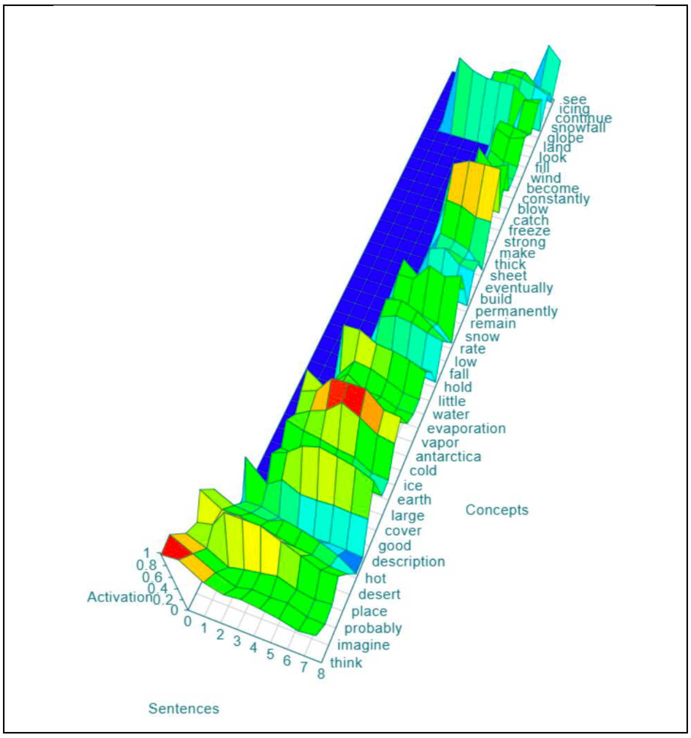

Figure 14. Activation scores provided by AMoC for a sample text.

4.7. Community Processing and Computer-Supported Collaborative Learning

Social Network Analysis (SNA) methods (Scott, 2017) can be used to model and analyze interactions between participants within a chat conversation or a community. Metrics like friendship relations, number of comments or thread discussions, can be used for this process. Computer-Supported Collaborative Learning (CSCL) (Stahl, Koschmann, & Suthers, 2006) environments facilitate collaborative learning through technology by using social interactions. Cohesion Network Analysis combines SNA with an in-depth discourse analysis to represent textual interactions between participants and generate sociograms.

The Community Processing and CSCL module of our *ReaderBench* framework has practical applications, both in educational and industry environments. For example, the discussions performed by students on the forum of a Massive Open Online Course (MOOC) platform can be analyzed, and their interactions can be modeled different interactive views (Dascalu et al., 2018; Sirbu et al., 2018). Figure 15 shows the interaction between participants from a math course using a clustered force layout. The nodes represent the participants, their size is direct proportional with the cumulated importance of their contributions and the colors reflect the three different clusters they were assigned: *central* (blue), *active* (green) and *peripheral* (orange). A similar approach can be applied to other communities, e.g., online blog communities in which different patterns can be observed (Sirbu et al., 2017).

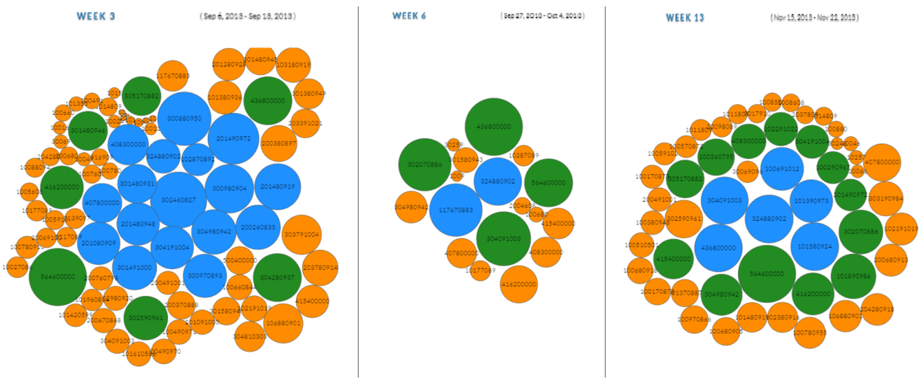

Figure 15. Weekly sociograms of the Community Processing service for a forum of an online Math course.

4. Educational Scenarios

We chose to design *ReaderBench* as a variety of independent services that can be used by anyone and can be easily integrated in other systems. This makes *ReaderBench* ready to be used in a large diversity of educational situations, as a Personalized Learning Environment.

From a teacher's point of view, the analysis of textual complexity for course materials may be a starting point in order to gauge the complexity of texts planned to be delivered to students. Second, teachers could use AMoC to test the likely understanding of students for specific parts of the material.

Regarding students' needs, the keyword extraction feature may be the most intensively used in order to create an overview of the main concepts from the read materials. Moreover, sentiment analyses would be also useful to assess the emotional charge of a literary text they were assigned to write.

In addition, both teachers and students would take benefit from the use of the CSCL analysis which displays valuable information, including students' degree of active participation and of collaboration during online discussions.

5. Conclusions and Future Work

This article introduces an updated version of our open-source NLP framework – *ReaderBench* – designed to support both students and tutors in multiple learning scenarios. Moreover, *ReaderBench* is designed to enhance the quality of learning processes conducted in multiple languages (English, French, Romanian, Dutch, Spanish, and Italian) and covering both individual and collaborative assessments. Our latest version (v4.1) includes: a) new features, APIs and visualizations, a new web interface for our presentation website, and the integration of new technologies. This version provides an increased performance due to the usage of spaCy for text-pre-processing instead of Stanford Core NLP, and AKKA for concurrent processing (Corlatescu, Paraschiv, Dascalu, Trausan-Matu, & Banica, 2018). Moreover, the new release provides greater modularity and ease of deployment through the usage of Artifactory and of Maven modules.

All services exposed within the updated *ReaderBench* website are presented, together with a brief description of the demos available online, usage examples and corresponding results. Applications in education and business fields were emphasized for each service. In contrast to the older version, a notable extension can be observed in the variety of provided services. This version also provides the opportunity for many more follow-up functionalities and improvements to be integrated in future releases, in accordance with the envisioned usage scenarios.

Acknowledgments

The work presented in this paper was funded by the README project "Interactive and Innovative application for evaluating the readability of texts in Romanian Language and for improving users' writing styles", contract no. 114/15.09.2017, MySMIS 2014 code 119286.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*(4-5), 993–1022.

Botarleanu, R.-M., Dascalu, M., Sirbu, M.-D., Crossley, S. A., & Trausan-Matu, S. (2018). ReadME – Generating Personalized Feedback for Essay Writing using the ReaderBench Framework. In H. Knoche, E. Popescu & A. Cartelli (Eds.), *3rd Int. Conf. on Smart Learning Ecosystems and Regional Development (SLERD 2018)* (pp. 133–145). Aalborg, Denmark: Springer.

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

Chen, Danqi, & Manning, Christopher. (2014). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 740–750). Doha, Qatar: ACL.

Corlatescu, D., Paraschiv, I. C., Dascalu, M., Trausan-Matu, S., & Banica, C. K. (2018). Concurrent Processing of Scientific Articles using Cohesion Network Analysis. In *17th Int. Conf. on Networking in Education and Research (RoEduNet)*. Cluj-Napoca, Romania: IEEE.

Cunningham, Hamish, Maynard, Diana, Bontcheva, Kalina, & Tablan, Valentin. (2002). A framework and graphical development environment for robust NLP tools and applications. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 168-175). Philadelphia, USA: ACL.

Dascalu, M. (2014). *Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence* (Vol. 534). Switzerland: Springer.

Dascalu, M., Crossley, S., McNamara, D .S., Dessus, P., & Trausan-Matu, S. (2018). Please ReaderBench this Text: A Multi-Dimensional Textual Complexity Assessment Framework. In S. Craig (Ed.), *Tutoring and Intelligent Tutoring Systems* (pp. 251–271). Hauppauge, NY, USA: Nova Science Publishers, Inc.

Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., & Nardy, A. (2014). Mining texts, learner productions and strategies with ReaderBench. In A. Peña-Ayala (Ed.), *Educational Data Mining: Applications and Trends* (pp. 345–377). Cham, Switzerland: Springer.

Dascalu, M., McNamara, D. S., Trausan-Matu, S., & Allen, L.K. (2018). Cohesion Network Analysis of CSCL Participation. *Behavior Research Methods, 50*(2), 604–619. doi: 10.3758/s13428-017-0888-4

Dascalu, M., Paraschiv, I. C., McNamara, D. S., & Trausan-Matu, S. (2018). Towards an Automated Model of Comprehension (AMoc). In *13th European Conference on Technology Enhanced Learning (EC-TEL 2018)* (pp. 427–436). Leeds, UK: Springer.

Dascalu, M., Sirbu, M.-D., Gutu-Robu, G., Ruseti, S., Crossley, S. A., & Trausan-Matu, S. (2018). Cohesion-Centered Analysis of Sociograms for Online Communities and Courses using ReaderBench. In *13th European Conference on Technology Enhanced Learning (EC-TEL 2018)*. Leeds, UK: Springer.

Dascalu, M., Trausan-Matu, S., McNamara, D.S., & Dessus, P. (2015). ReaderBench – Automated Evaluation of Collaboration based on Cohesion and Dialogism. *International Journal of Computer-Supported Collaborative Learning, 10*(4).

Fellbaum, C. (2005). WordNet(s). In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (2nd ed., Vol. 13, pp. 665–670). Oxford: Elsevier.

Gutu, G., Dascalu, M., Trausan-Matu, S., & Dessus, P. (2016). ReaderBench goes Online: A Comprehension-Centered Framework for Educational Purposes. In A. Iftene & J. Vanderdonckt (Eds.), *Romanian Conference on Human-Computer Interaction (RoCHI 2016)* (pp. 95–102). Iasi, Romania: MATRIX ROM.

Gutu, G., Paraschiv, I. C., Dascalu, M., Cristian, G., Trausan-Matu, S., & Lepoivre, O. (2018). Analyzing and Providing Comprehensive Feedback for French CVs with ReaderBench. *Scientific Bulletin, University Politehnica of Bucharest, Series C, 80*(2), 17–28.

Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*. Seattle, WA: ACM.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*(2), 211–240.

Lasswell, H. D., & Namenwirth, J. Z. (1969). *The Lasswell Value Dictionary*. New Haven: Yale University Press.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In *52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 55–60). Baltimore, MA: ACL.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representation in Vector Space. In *Workshop at ICLR*. Scottsdale, AZ.

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

Monnier, Catherine, & Syssau, Arielle. (2014). Affective norms for french words (FAN). *Behavior Research Methods, 46*(4), 1128-1137. doi: 10.3758/s13428-013-0431-1

Moors, Agnes, De Houwer, Jan, Hermans, Dirk, Wanmaker, Sabine, van Schie, Kevin, Van Harmelen, Anne-Laura, De Schryver, Maarten, De Winne, Jeffrey, & Brysbaert, Marc. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods, 45*(1), 169-177. doi: 10.3758/s13428-012-0243-8

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

Scott, John. (2017). *Social network analysis*. London, UK: Sage.

Sirbu, M.-D., Botarleanu, R., Dascalu, M., Crossley, S., & Trausan-Matu, S. (2018). ReadME – Enhancing Automated Writing Evaluation. In *18th Int. Conf. on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA 2018)* (pp. 281–285). Varna, Bulgaria: Springer.

Sirbu, M.-D., Dascalu, M., Crossley, S. A., McNamara, D. S., Barnes, T., Lynch, C. F., & Trausan-Matu, S. (2018). Exploring Online Course Sociograms using Cohesion Network Analysis. In C. P. Rosé, R. Martínez-Maldonado, U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren & B. d. Boulay (Eds.), *19th Int. Conf. on Artificial Intelligence in Education (AIED 2018), Part II* (pp. 337–342). London, UK: Springer.

Sirbu, M.-D., Panaite, M., Secui, A., Dascalu, M., Nistor, N., & Trausan-Matu, S. (2017). ReaderBench: Building Comprehensive Sociograms of Online Communities. In *9th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2017)*. Timisoara, Romania: IEEE.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge, UK: Cambridge University Press.

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