Analyzing and Providing Comprehensive Feedback for French CVS with Readerbench

ANALYZING AND PROVIDING COMPREHENSIVE FEEDBACK FOR FRENCH CVS WITH READERBENCH

Gabriel GUȚU[1], Ionuț Cristian PARASCHIV[2], Mihai DASCĂLU[3], Gabriel CRISTIAN[4], Ștefan TRĂUȘAN-MATU[5], Olivier LEPOIVRE[6]

In their everyday activities, recruiters are faced with the difficult task of analyzing and judging the quality of a wide range of CVs. Both the content quality and the visual hues, such as colors and their overall structure, need to be considered. This article enhances previous researches with a larger dataset, refined indices, and a more advanced technique of parsing the input documents. After applying various processing techniques from ReaderBench, an advanced Natural Language Processing framework, on a manually annotated dataset of 96 positive and negative French CVs, several writing indices were determined and filtered by leveraging statistical analyses. In addition, our experiment introduces a web application in which users can submit, gather an evaluation, and acquire valuable feedback on their CV.

Keywords: CV analysis, CV assessment, text cohesion, textual complexity, Natural Language Processing

1. Introduction

Automated CV (Curriculum Vitae) analysis is a method that can enhance the recruiting process by categorizing and scoring resumes through various metrics. This task can potentially decrease the employment turnover which includes finding candidates, gathering, analyzing and filtering lists of CVs, followed by interviews and the selection of the most adequate candidate(s). Thus, the selection process can be facilitated by analyzing the CVs which have the highest chance to contain more relevant and better structured information, and by

[1] PhD student, Dept. of Computer Science, University POLITEHNICA of Bucharest, Romania, e-mail: gabriel.gutu@cs.pub.ro
[2] PhD student, Dept. of Computer Science, University POLITEHNICA of Bucharest, Romania, e-mail: ionut.paraschiv@cs.pub.ro
[3] Reader, Dept. of Computer Science, University POLITEHNICA of Bucharest, Romania, e-mail: mihai.dascalu@cs.pub.ro
[4] Master degree student, Dept. of Computer Science, University POLITEHNICA of Bucharest, Romania, e-mail: gabriel.cristian0310@stud.acs.upb.ro
[5] Prof., Dept. Dept. of Computer Science, University POLITEHNICA of Bucharest, Romania, e-mail: stefan.trausan@cs.pub.ro
[6] Responsable du Domaine Digital, Groupe Randstad France, Saint Denis, France, e-mail: olivier.lepoivre@randstad.fr

grouping CVs into *positive* and *negative* categories for specific criteria. The final decision of selecting candidates is performed by the recruiter; nonetheless, human-aided decision support systems are useful especially when doing redundant work that requires focus for large amounts of time.

The presented experiment is used within a serious game (Job Quest) that allows applicants to automatically asses the quality of their business or management-oriented CV written in French language. The game is aimed at providing feedback to learners in order to enhance their CV for a required management position. The training dataset of CVs reflects the company's specific recruitment requirements and enables potential candidates to adapt their CVs in accordance.

This article extends the research performed by Gutu et al. [9], which describes a similar approach for analyzing and categorizing CVs based on a smaller corpus, with no feedback provided, and no overall scores assigned. The current research uses a number of 96 CVs pre-tagged with a positive or a negative class on two different aspects: *visual appearance* and *textual content*. For each CV from the corpus, several visual attributes (e.g., font size, text color) and over 400 textual complexity indices (e.g., number of characters, commas) have been computed and were introduced within a statistical analysis that determines the most adequate ones to be used in categorizing new CVs as *positive* or *negative*. Moreover, the selected metrics are used within a web application in which users are able to upload and assess their own CV, written in French language and designed for business purposes.

Various online platforms exist that enable users to upload and check the adequacy of their CVs (see ***Table 1***). However, most of the existing systems have been developed in the context of linking employers with potential employees and rely on closed algorithms and heuristics. In this context, we have opted to build our own web application on top of *ReaderBench* [2; 3; 5], the framework used for computing the textual indices presented in the upcoming section. Subsequently, the research continues with the description of the new corpus and of the results, followed by conclusions.

Table 1

Existing CV Analyzers

Live Career livecareer.com/resume-check	Platform focusing on the interview process for employers and employees; verifies verbs, font, grammar, etc.
Employment Boost employmentboost.com/free-resume-review-evaluation/	Manual CV assessment performed by a professional reviewer.
The Ladders theladders.com/resume-reviewer	Platform for connecting employers with potential employees that uses a closed algorithm for scoring CVs.

2. The ReaderBench Framework

ReaderBench is an open source framework containing a fully functional Natural Language Processing pipeline capable of annotating raw corpora using various semantic models alongside text preprocessing, part of speech tagging, syntactic dependencies, textual complexity indices and sentiment analysis. *ReaderBench* was employed in other researches where its textual complexity indices were used to compute properties such as readability, syntax, discourse structure or semantics [2; 4]. Semantic models are pre-trained on large textual corpora which are used to tag unstructured documents for computing semantic distances between their vector representations. Semantic processing usually considers documents as a *bag of words*, meaning that the order in which words appear is not important, only their count. Within the current experiment, the CVs have been annotated using 3 different semantic models built in *ReaderBench*. First, Latent Semantic Analysis (LSA) [10] creates a term-document matrix which stores the normalized number of occurrences of each word in every document and applies a Singular Value Decomposition, thus resulting the document and word vector representations. Second, Latent Dirichlet Allocation (LDA) [1] uses a probabilistic model to map words and documents as distributions of latent topics. Third, word2vec [11] is a novel semantic model relying on word embeddings that has a faster training time. By combining these three models, semantic distances can be estimated with higher precision.

In addition, *ReaderBench* includes lists of words having similar lexical, sociological or psychological meanings, as well as lists of word valence scores. For French, the framework integrates *Affective norms for French words* (FAN) [12] and *Linguistic Inquiry and Word Count* (LIWC) [13]. FAN contains 1,034 words with valences between -3.8 and 3.8; thus, words were tagged as *negative* below -1 and as *positive* over 1, considering neutral the words with values between -1 and 1. LIWC classifies more than 4,500 words or stems based on various semantic aspects. Several categories are of particular interest for this experiment as they express strong sentiments such as anger, anxiety or aggression, thus potentially leading to a higher classification accuracy of CVs [15].

3. Dataset

The experiments were conducted on a collection of 96 CVs written in French language that also includes the 52 CVs used in the previous experiment performed by Gutu, Dascalu, Trausan-Matu and Lepoivre [9]. The CVs were provided by the French division of Randstad, a global leader in the Human Resources (HR) services industry. They were curated by preserving only the CVs that had feedback scores provided and that were not scanned documents. As an overview, the considered CVs tend to be quite short (one, maximum two pages)

and state short descriptions together with financial figures for previous work positions.

For scoring the CVs, the HR recruiter was asked to provide a score, either 1 as positive or -1 as negative, for three dimensions. Since only one rater could be internally appointed due to the specificity of the task, no inter-rater agreement score could be computed, and we had to rely on the sole expert's opinion. The first dimension covered the visual aspects of the CVs, namely the *visual score*. The second dimension was related to the textual contents of the CVs, later on called the *content score*. The last dimension of the manual annotation process referred to the overall score for the CVs, i.e. its *global score*.

Of the proposed collection of more than 100 CVs, we imposed a balance between the number of positive versus negative ones regarding the global score by removing part of them; in the end, 96 CVs were selected. Table 2 shows the distribution of positive versus negative CVs with regards to each criterion.

Table 2

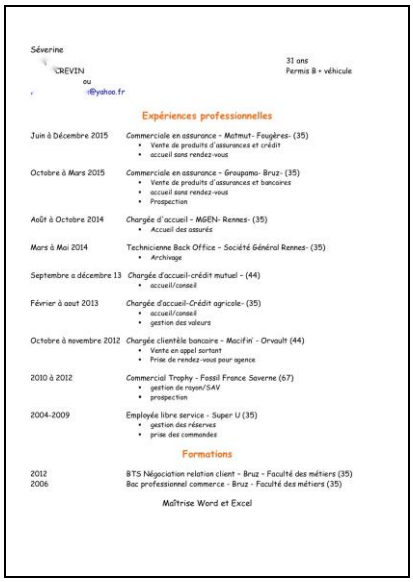
Number of positive versus negative CVs for each criterion (total = 96 CVs)

Criterion	No. positive CVs	No. negative CVs
Global score	57 (59.38%)	39 (40.62%)
Visual score	56 (58.33%)	40 (41.67%)
Content score	55 (57.29%)	41 (42.71%)

The average file size of a CV was 0.25 MB (*SD* = 0.21MB, *min* = 0.01MB, *max* = 1.22MB). Differentiated by the global criterion, positive CVs sized about 0.28MB (*SD* = 0.19MB, *min* = 0.01MB, *max* = 0.81MB), while negative CVs were in general smaller, averaging about 0.20MB (*SD* = 0.22MB, *min* = 0.01MB, *max* = 1.22MB).

Fig. 1 depicts samples of CVs labelled as “positive”, respectively as “negative”, for all the three dimensions: visual, content-related aspects and global perspective. The positive CV contains qualitative information including well-structured text into categories of competence, description of the previous jobs together with the period of times, and detailed information about the candidate’s education. The visual aspects are also positive because of the CV’s layout, delimitation of sections based on colors, and the usage of different types of fonts that allow the understanding of important elements with ease.

a) CV labelled as "positive" for global, visual and content-related criteria



b) CV labelled as "negative" for global, visual and content-related criteria

Fig. 1. Comparison of CVs labelled as "positive" and "negative" for the three criteria

The CV annotated as negative for all three criteria uses mostly the same font, thus not delimiting sections and its content is quite scarce. We need to emphasize that the two CVs were chosen to emphasize rather extreme good and bad scenarios; our collection contained also CVs that were annotated as positive on one criterion, while negative on the others and vice-versa, based on the annotator's decision.

4. Method

The semantic models used in this experiment require for training a large collection of documents. The *Le Monde* corpus with more than 22 million words[7] was used for training the LSA, LDA and word2vec models. The corpus consists of articles extracted from the French *Le Monde* newspaper (1993 edition) and covers both general and specific terms, gathered from politics, culture, sport, science, society or finance. This allows *ReaderBench* to conceptualize a large number of content words (i.e., nouns, verbs, adjectives and adverbs) encountered in the processed CVs. Starting from the three dimensions defined for the manual annotation process, the visual aspects covered facets like document clearness and spacing, colors, images and layout, the utilized font types and the corresponding sizes. The coverage of bold, italic and both bold and italic characters represented

[7] http://lsa.colorado.edu/spaces.html

another factor in scoring a CV as positive or negative with regards to its visual aspects.

The textual content aspects covered word complexity, spelling, and the consideration of words from pre-imposed categories (e.g., "*anger*", "*sadness*", "*inclusion*", or "*perception*"). The average number of paragraphs per page, together with the average number of sentences per paragraph and the number of words within them, were also considered for the content score. Positive, negative and neutral words, together with indices regarding textual complexity and cohesion between units of texts were also counted as important factors into scoring the content of the CV. The global score was graded by considering all previous indices in order to assess the overall perception of a CV's quality.

Three statistical analyses were carried out to assess each dimension. More than 400 indices were calculated, out of which about 5% were used for the visual score, while 95% for the content score. The global analysis took into consideration all existing indices. The visual indices were computed using the Apache PDFBox library[8] that extracts text from PDF documents, as well as formatting information including fonts, corresponding sizes and colors, together with image statistics.

Compared to a previous experiment [9], our approach uses a more adequate identification of paragraphs. The conversion of texts extracted from PDF files using tools like Apache PDFBox does not always reflect the structure that a human would consider. This is mainly caused by the various sizes of spacing between different paragraphs, between rows belonging to the same paragraph, between section titles and their content, and even between columns if the text is displayed in such a design. While the previous experiment used an approach in which single end-of-lines were removed and multiple ones were replaced with one paragraph, we now rely on a more advanced approach. The functionality to determine sequences of types of fonts (size, font type, numbering, etc.) enabled us to build a hierarchical structure in which elements of the same formatting (e.g., heading 1, heading 2, paragraph, etc.) are assigned at the same level within the resulting tree. The entire text that follows a certain level is assigned to a singular paragraph, thus drastically reducing the number of one-line paragraphs that were previously encountered. Therefore, the average number of paragraphs was decreased by 49.25% (from 34.48 to 17.50), while the standard deviation decreased by 44.38% (from 17.50 to 9.95).

In order to validate the quality of the structures determined by our new approach, we computed the difference between the number of paragraphs found with the two approaches and emphasized the differences outside one standard deviation from the average; 24% of the CVs satisfied this criterion, thus a group of experts were asked to provide a number of expected paragraphs. The difference

[8] https://pdfbox.apache.org

between the algorithm's paragraphs and the average of experts' annotations lead to 3.85 paragraphs (*SD* = 10.92) for the new algorithm compared to an average of 29.07 paragraphs (*SD* = 25.20) for the old approach. Thus, an improvement of 86.76% was observed in the new algorithm's ability to detect paragraphs.

5. Results

The afore-mentioned indices were computed for the entire collection of CVs and only the ones showing normality were preserved. For the indices that showed multicollinearity (estimated as pairwise correlations with $r > .70$), the one with the stronger effect was preserved for further analyses. Three multivariate analyses of variance (MANOVA) [6] were performed to determine the effect of each remaining index for the considered dimensions: visual, content-related and global quality of the CV. Table 3 shows the indices identified as statistically significant for the visual criteria ($p < 0.05$), ordered in descending order of their effect size (F). Table 4 depicts the equivalent analysis results on the content-related criterion, while Table 5 shows results from the global perspective. Our CVs exhibited a lower minimum font size for positive CVs than for the negative ones. Moreover, the average number of paragraphs per page was much higher for positive CVs. The content aspects showed multiple significant indices, out of which higher values for positive CVs were obtained for: average number of syntactic dependencies per sentence, average number of words within the "Insight" list (expressing judgement and cognitive processes) per sentence, average number of prepositions per sentence, sentence standard deviation in terms of words and unique words alongside a few others.

Afterwards, three stepwise Discriminant Function Analyses (DFAs) were performed. The first DFA, performed on the visual criterion, retained one predictor as significant: *minimum font size* (Wilks' $\lambda = .940$, $\chi^2(df = 1) = 5.833$, $p = .016$). The DFA correctly allocated 60 (30 + 30) CVs with an accuracy of 62.5% using a leave-one-out cross-validation. The second criterion, the content-related aspects, lead to a DFA model that retained one variable (*Average number of syntactic dependencies between a verb and one of its accusative objects per paragraph*) as significant with Wilks' $\lambda = .900$, $\chi^2(df = 1) = 9.876$, $p = .002$. The DFA correctly allocated 58 CVs (29 + 29) with an accuracy of 60.4% using a leave-one-out cross-validation. The last dimension, global quality, retained one variable (*minimum font size*) as significant predictor, similar to the visual dimension. The resulted DFA model was able to significantly differentiate between CVs on the global score, Wilks' $\lambda = .941$, $\chi^2(df = 1) = 5.649$, $p = .017$. The DFA correctly allocated 57 CVs (29 + 28) with an accuracy of 59.4% using a leave-one-out cross-validation. Afterwards, the resulting Fischer coefficients were

used to score CVs as "positive" or "negative" based on the predictive indices for each dimension.

Table 3.

Test of Between-Subjects Effects for Predictive Indices of the Visual Criteria

Index	***M (SD) positive***	***M (SD) negative***	***F***	***p***	***Partial η^2***
Minimum font size	5.84 (3.64)	7.72 (3.75)	6.051	.016	.060
Average number of paragraphs per page	16.89 (9.15)	12.55 (7.67)	5.986	.016	.060

Table 4.

Test of Between-Subjects Effects for Predictive Indices of the Content-related Criteria

Index	***M (SD) positive***	***M (SD) negative***	***F***	***p***	***Partial η^2***
Average number of syntactic dependencies between a verb and one of its accusative objects per sentence	0.61 (0.43)	0.37 (0.24)	10.472	.002	.100
Average number of words labelled as "insight" per sentence (LIWC)	0.51 (0.32)	0.35 (0.18)	8.730	.004	.085
Average number of prepositions per sentence	0.89 (0.55)	0.62 (0.31)	7.854	.006	.077
Sentence standard deviation in terms of unique content words	4.59 (2.46)	3.46 (1.49)	6.843	.010	.068
Average number of nouns per sentence	5.77 (2.71)	4.51 (1.70)	6.811	.011	.068
Average number of words labelled as "money" per document (LIWC)	17.38 (10.37)	12.71 (7.53)	5.978	.016	.060
Average number of syntactic dependencies between a verb and one of its accusative objects per paragraph	0.68 (0.54)	0.44 (0.46)	5.577	.020	.056
Standard deviation of sentence relevance scores	2.34 (1.37)	1.79 (0.90)	5.006	.028	.051
Average number of syntactic dependencies containing proper names per paragraph	1.01 (0.59)	0.77 (0.49)	4.645	.034	.047
Average number of syntactic dependencies from a noun to the head of an appositive NP per sentence	1.50 (0.92)	1.16 (0.61)	4.388	.039	.045
Average number of words labelled as "relativity" per sentence (LIWC)	1.19 (0.67)	0.94 (0.48)	4.175	.044	.043
Average number of words labelled as "friends" per document (LIWC)	4.44 (3.37)	3.17 (2.45)	4.155	.044	.042
Average sentence length	41.52 (21.95)	33.92 (13.59)	4.100	.046	.042
Average number of unique prepositions per	0.70	0.54	4.092	.046	.042

Index	*M (SD) positive*	*M (SD) negative*	*F*	*p*	*Partial η^2*
paragraph	(0.45)	(0.29)			
Average number of words labelled as "tentative" per document (LIWC)	1.67 (1.56)	1.05 (1.40)	4.093	.046	.042
Average sentence relevance score	2.18 (1.18)	1.76 (0.75)	4.023	.048	.041

Table 5. Test of Between-Subjects Effects for Predictive Indices of the Global Criteria

Index	*M (SD) positive*	*M (SD) negative*	*F*	*p*	*Partial η^2*
Minimum font size	5.87 (3.72)	7.72 (3.65)	5.854	.017	.059
Average number of paragraphs per page	16.77 (9.13)	12.61 (7.73)	5.450	.022	.055

6. The Online Tool

A web client corresponding to our CV analysis tool is available freely online in the *ReaderBench* website[9] [8]. The tool allows individuals to upload their French management-oriented CV in PDF format and gather valuable feedback from the previously described statistical analyses. The usage of specific keywords as input data allows an employer to grasp at a glance whether a candidate covered concepts required for their position. The option of ignoring specific words helps removing irrelevant information, i.e. names of months, days, streets, or any other words that are not considered relevant by the employer. This list of words complements the process of removing stop-words that are automatically deleted by the *ReaderBench* framework.

Other parameters allow the user to test different scenarios by switching between different pre-trained models like LSA, LDA and word2vec. Other options refer to enabling part of speech tagging on one hand, which increases the quality of results as our tool is capable of differentiating between parts of speech, and the enabling of dialogism computation, on the other hand; this option is a more advanced feature of *ReaderBench* that is used to determine voices and lexical chains, a functionality not relevant for the assessment of CVs. The threshold for semantic similarity imposes a minimum strength of the links between words in the generated concept map, i.e. a network graph of the most important keywords from the text and the corresponding semantic relations exceeding the imposed threshold.

Our CV analysis tool provides personalized recommendations based on the feedback score for each of the three dimensions. Three scores are obtained,

[9] http://readerbench.com/demo/cv

one for each criterion, together with their corresponding messages. The feedback obtained for the CVs depicted in **Fig. 1** includes the following messages:

- *"Votre CV est globalement satisfaisant. Toutes nos félicitations!"* (*en*: "Your CV is overall more than satisfying. Congratulations!"; Global score: 1)
- *"Le visuel de votre CV est parfait! Toutes nos félicitations!"* (*en*: "Your CV is perfect from a visual point of view. Congratulations!"; Visual score: 1)
- *"Vous devriez améliorer le contenu de votre CV. Il manque de pertinence."* (*en*: "You need to enhance the contents of your CV as it lacks pertinence"; Content-related score: -1)

As the training dataset was quite small and accuracy was low, thus denoting inconsistencies in the manual assessment, we propose this tool as a promoter for the automated analysis of CVs, a field currently in an early stage of development.

7. Conclusions

This article presents novel research with regards to assessing the quality of CVs. Our method uses a more adequate approach to segment the original text into cohesive paragraphs, significantly closer to human annotations, and our experiments were performed on a larger dataset. Visual and content-related indices were used to categorize CVs as positive or negative for each of the three dimensions: visual, content-related and global perspective. Having the ability to easily measure the adequacy of CVs can potentially decrease the time of analyzing candidates and finding the ones which are the most suitable for an interview.

The small size of our dataset was induced by the following limitations: first, CVs include sensitive data for applicants and require their explicit consent; second, the consideration of specific French CVs related to financial positions limits the number of potential entries; third, CVs need to be manually evaluated by a recruiter designated by the company. The subjectivity of the annotator's scores is reduced by the specificity of the scenario involved in the experiment, which refers to finding the most suitable candidates for jobs that imply management / financial positions. However, the presented results might not fit other use cases; for example, an Information Technology and Communications company might consider more important known programming languages, professional certifications or the performance of an algorithm designed by the candidate; thus, the relevance scores are potentially different between different domains. Nevertheless, our tool can be easily adapted for usage within other

companies or domains by training it on a specific dataset of CVs and by enforcing new keywords to be covered.

As future work, we envision that CVs can be matched with corresponding job descriptions using semantic similarity which can help even further to find suitable candidates, alongside their CV's quality score. Even more, an equivalent analysis performed on a collection of CVs written in English might highlight similarities and potential differences between the two languages. In addition, the integration of a crawling algorithm of online communities built over professional social networks like LinkedIn using tools as BlogCrawl [14] might allow employees to build a network of experienced representatives from a domain. Moreover, complementary experiments could consider the integration of additional characteristics through the extraction of online texts as individuals frequently produce content on multiple social platforms and websites. Thus, specific attributes and corresponding values could be automatically extracted [7] and used to describe the personal characteristics, personality and writing style traits of each candidate.

Acknowledgement

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

R E F E R E N C E S

[1] Blei, D.M., Ng, A.Y., and Jordan, M.I., 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research 3, 4-5, 993–1022.

[2] Dascalu, M., 2014. Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence. Springer, Cham, Switzerland.

[3] Dascalu, M., Dessus, P., Trausan-Matu, S., Bianco, M., and Nardy, A., 2013. ReaderBench, an environment for analyzing text complexity and reading strategies. In Proceedings of the 16th Int. Conf. on Artificial Intelligence in Education (AIED 2013) (Memphis, USAYear), Springer, 379–388.

[4] Dascalu, M., Gutu, G., Ruseti, S., Paraschiv, I.C., Dessus, P., McNamara, D.S., Crossley, S., and Trausan-Matu, S., 2017. ReaderBench: A Multi-Lingual Framework for Analyzing Text Complexity. In Proceedings of the 12th European Conference on Technology Enhanced Learning (EC-TEL 2017) (Tallinn, EstoniaYear), Springer, 495–499.

[5] Dascalu, M., Stavarache, L.L., Dessus, P., Trausan-Matu, S., McNamara, D.S., and Bianco, M., 2015. ReaderBench: An Integrated Cohesion-Centered Framework. In Proceedings of the 10th European Conf. on Technology Enhanced Learning (Toledo, SpainYear), Springer, 505–508.

[6] Garson, G.D., 2015. Multivariate GLM, MANOVA, and MANCOVA. Statistical Associates Publishing, Asheboro, NC.

[7] Ghecenco, A., Rebedea, T., and Chiru, C., 2017. Extraction of Attributes and Values From Online Texts. Scientific Bulletin, University Politehnica of Bucharest, Series C 79, 1, 67-82.

[8] Gutu, G., Dascalu, M., Trausan-Matu, S., and Dessus, P., 2016. ReaderBench goes Online: A Comprehension-Centered Framework for Educational Purposes. In Proceedings of the Romanian Conference on Human-Computer Interaction (RoCHI 2016) (Iasi, RomaniaYear), MATRIX ROM, 95–102.

[9] Gutu, G., Dascalu, M., Trausan-Matu, S., and Lepoivre, O., 2017. How Adequate is your CV? Analyzing French CVs with ReaderBench. In Proceedings of the 3rd Int. Workshop on Design and Spontaneity in Computer-Supported Collaborative Learning (DS-CSCL-2017), in conjunction with the 21th Int. Conf. on Control Systems and Computer Science (CSCS21) (Bucharest, RomaniaYear), IEEE, 559–565.

[10] Landauer, T.K., Foltz, P.W., and Laham, D., 1998. An introduction to Latent Semantic Analysis. Discourse Processes 25, 2/3, 259–284.

[11] Mikolov, T., Chen, K., Corrado, G., and Dean, J., 2013. Efficient Estimation of Word Representation in Vector Space. In Proceedings of the Workshop at ICLR (Scottsdale, AZYear).

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

[13] Piolat, A., Booth, R.J., Chung, C.K., Davids, M., and Pennebaker, J.W., 2011. La version française du LIWC : modalités de construction et exemples d'application. Psychologie française 56, 145–159.

[14] Stavarache, L.L., Balint, M., Dascalu, M., Trausan-Matu, S., and Nistor, N., 2017. BlogCrawl: Customized Crawling of Online Communities. University Politehnica of Bucharest Scientific Bulletin Series C-Electrical Engineering and Computer Science 79, 2, 3–14.

[15] Tausczik, Y.R., & Pennebaker, J. W., 2010. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology 29, 1, 24-54.