Word Clouds as a Learning Analytic Tool for the Cooperative e-Learning Platform NeuroK

Fernando Calle-Alonso[1], Vicente Botón-Fernández[2], Jesús M. Sánchez-Gómez[2], Miguel A. Vega-Rodríguez[3], Carlos J. Pérez[4] and Daniel de la Mata[1]

[1]*Research & Development Department, ASPgems SL, Spain*
[2]*Cátedra ASPgems, University of Extremadura, Spain*
[3]*Department of Technologies of Computers & Communications, University of Extremadura, Spain*
[4]*Department of Mathematics, University of Extremadura, Spain*

Keywords: Word Cloud, Tag Cloud, e-learning, Online Learning, NeuroK, Evaluation, Engagement.

Abstract: Word cloud or tag cloud is very popular these days. It is a tool used to display text data summarization in a visual way very easy to understand. However, it has not been extensively used in teaching, especially in e-learning, where it would make a differential advantage. This research presents the definition and implementation of a word cloud tool in a social network-based e-learning platform (NeuroK), which is based on the principles of neurodidactics. The different features developed and the results are shown. Several options to compare word clouds from students and teachers allow the teacher to follow the development of the course, and they provide him more information to facilitate the evaluation process.

1 INTRODUCTION

Nowadays the use of e-learning platforms and learning management systems is very common for educational institutions and companies. Although digital education is growing continuously some problems still should be resolved to continue its development, for example, the difficulty to evaluate students (dos Santos and Favero, 2015), the prevention of dropping out (Yukselturk, 2014), the redesign of educator roles (Adams et al., 2017), understanding automatically the natural language (Aeiad and Meziane, 2016), the flaws in the accreditation process of distance and flexible learning programs (Reeves, 2003) or blending formal and informal learning (Czerkawski, 2016).

Calle-Alonso et al., 2017 developed a new e-learning platform based on the principles of neurodidactics called NeuroK. This platform tries to solve several of the current problems affecting online learning environments.

Learning analytics (Baker and Inventado, 2014) allows to track students' performance by using the data obtained from their activities (connections, comments, evaluations, documents shared, favorites...). With learning analytics and visualizations, real-time analysis of the course could be performed and future-tense adapting actions could be carried out to anticipate the course drifting in the wrong direction (see, e.g., Nevado-Maestre et al. 2017).

Word clouds or tag clouds can be used as learning analytics tools. They are visual representations of a group of words used by the participants, and based on their frequency. These kinds of clouds give greater prominence to the words appearing more frequently and reflect on all the information from within the texts written by students and teacher in a course. By investigating the patterns of words or phrases, or lack thereof, in textual student responses, instructors can evaluate if students, as a whole, have grasped or missed key concepts or have made common mistakes (De Paolo and Wilkinson, 2014). Word clouds belong to Natural Language Processing (NLP) field (Heimerl et al., 2014). They are very easy to understand and they can be included into any class, subject and age. Although word clouds are very powerful, they are not used very much in education (nor in online education of course), but the use of word clouds could offer a lot of benefits in e-learning for both, teachers and students.

Nickell (2012) shows how word clouds work and test them in mathematics classrooms. He showed how

Calle-Alonso, F., Botón-Fernández, V., Sánchez-Gómez, J., Vega-Rodríguez, M., Pérez, C. and de la Mata, D.
Word Clouds as a Learning Analytic Tool for the Cooperative e-Learning Platform NeuroK.
DOI: 10.5220/0006816505080513
In *Proceedings of the 10th International Conference on Computer Supported Education (CSEDU 2018)*, pages 508-513
ISBN: 978-989-758-291-2

to promote engagement in online learning. Perry (2012) presents a use case to teach speaking style with word clouds. Miley and Read (2011) used Wordle (see Feinberg, 2014) to introduce Word Clouds in a classroom routine, obtaining remarkable results, validated by the student's opinions. Finally, also with Wordle, McNaught and Lam (2010) applied word cloud tool for content analysis in research. These are some interesting results of the application of Word Clouds in teaching and research.

NeuroK is meant to be a social network based learning platform, and the learning process is supported by discussions. These interactions motivate students and engage them. Word cloud tool is demonstrated to be engaging for the students, and to motivate them. Kaptein et al. (2010) and Behar-Horenstein and Niu (2011) established that visual representations of students texts with a word cloud improved both critical thinking and engagement, especially in online discussions. It was tested by deNoyelles and Reyes-Foster (2015), revealing that students using word clouds reported moderately higher scores on critical thinking and engagement, as well as peer interaction.

In the next sections, NeuroK Word Cloud tool is presented and the features and results are shown with examples.

2 IMPLEMENTATION

Today, word clouds have been shown to be a beneficial tool in many educational environments (Miley and Read, 2011). They provide a fast visual representation of the information contained in texts. In this work, word clouds are built upon a cooperative e-learning platform, Neurok, to provide an overview of the main concepts used by the students in their communications. Thereby, teachers can analyze what is being talked about in a course and make decisions on strategy to guide the learning process through the texts the students write. And since it is a learning methodology based on the content of communications, the participation and activity of students is promoted and supported.

The tool under consideration consists of three different word clouds: the target cloud, with the concepts that the teacher would like their students to talk about; the real cloud, with the concepts most commonly used by the students; and the mixture of both, with the target concepts according to their use by the students. These three word clouds provide a summary at both topic and course level. Several metrics to measure the concordance between the target cloud and the real cloud have also been implemented.

2.1 Word Cloud

Word clouds represent the concepts most frequently used by students. In our word cloud representation, the bigger and more colourful a concept is, the larger the number of times it is referred to.

The data obtained from NeuroK to build the word clouds and to compute the concordance are mainly:

- Remarks: these are opinions and statements that a student publishes in a course or topic.
- Comments: these are the replies that other students make of a certain remark.
- Rates: these are comments that a student makes to justify their assessment of a given remark.
- Delivery documents: these are the documents that a student encloses in a delivery. Valid extensions are .doc, .docx, .pdf, .pptx and .txt.

Each time any of these contents are published in a course or topic, the corresponding text is automatically analysed. First of all, words are extracted from the text and those that do not provide useful information, such as e-mails, URLs, numbers, acronyms… are removed from the analysis process. Once the words have been extracted, they are analysed one by one checking they are not stop words and obtaining their stem. Stop words are the most common words in a language (prepositions, articles, adverbs, conjunctions, pronouns and some verbs), whereas a stem is the root form of a word. Stop words are filtered because they are not so relevant for natural language processing purposes since they occur frequently in a language and bring little semantic value to the content. At present, there is no single universal list of stop words available, so we have built our own list. Furthermore, a stemming process based on Porter Stemming Algorithm (Porter, 1997) is used for getting the stem words. Other versions of the algorithm have been included to support more languages, such as Spanish (Porter, 2001).

Finally, both the analyzed word and those with which it is related are mapped to the same stem, and one of them is selected to represent them all. This avoids repetition of words with the same meaning in the resultant cloud. When choosing the most representative word, several criteria are taken into account: infinitive verb, shorter noun, etc. Words frequency is counted by stem and only those with the highest frequency are shown in the cloud.

2.2 Target Cloud

The target cloud is a special type of cloud that represents the ideal word cloud from the teacher's point of view. In this cloud, the teacher can add new words and specify the weight of each word. By default, this weight is set proportionally to the number of target words. To add new words, the teacher can either type in a new one or select one from the list of words that the students have used. Once the target cloud is defined, the teacher will be able to compare it with the course word cloud and analyse the concordance.

2.3 Interface

For the representation purpose, *jQCloud* plugin has been used. This plugin has some advantages such as:

- Dynamic lightweight and customizable tag cloud.
- Cloud shaped appearance.
- Vertical, elliptic and rectangular clouds support.
- Custom tag's links, styles and weights.

The teacher and students can interact with the word cloud by doing any of the following actions:

- Choosing the course or topic they want to be represented in the cloud.
- Filtering data by user: students and teacher/s, only students or a specific student.
- Selecting the size of the cloud: 10 words, 25 words or 50 words.
- Comparing the current cloud with the target cloud.

However, there are differences between students and teachers in terms of permissions. On the one hand, students can see highlighted in the mixture of clouds those words they are talking about and that the teacher wants them to use, but they cannot see the target words they are not using yet. On the other hand, the teacher can see all types of words: the target words the students are talking about, the remaining target words that have not been used, and the words the students are using but are not part of the target cloud. In addition, teachers are the only ones with enough permissions to edit the target cloud.

3 A LEARNING ANALYTIC TOOL

In this section, a typical scenario that covers the concepts previously mentioned has been described in order to better understand the system operation. We have tested NeuroK through an existing topic called "Machine Learning" to obtain a real dataset. There are ten students enrolled in the course and two teachers. In order to carry out the tag clouds service, all the information from remarks, comments, delivery documents and rates are saved into the NeuroK database. This information is registered during a period of one month. Once the dataset is ready, it is time to navigate to the "Tag Cloud" view and set up the filters. In this case, we establish the following settings: Machine Learning topic, only students and 25 most relevant words. The rest of the filters remain as default.

After running our tag cloud approach over the previous dataset, it generates the word cloud that appears in Fig. 1. This cloud consists of the 25 most frequently used words by students of the Machine Learning topic. In order to evaluate this word cloud it is necessary to define the target cloud. To do this, the teacher can click on the button "Manage target cloud" and customize its own target cloud.

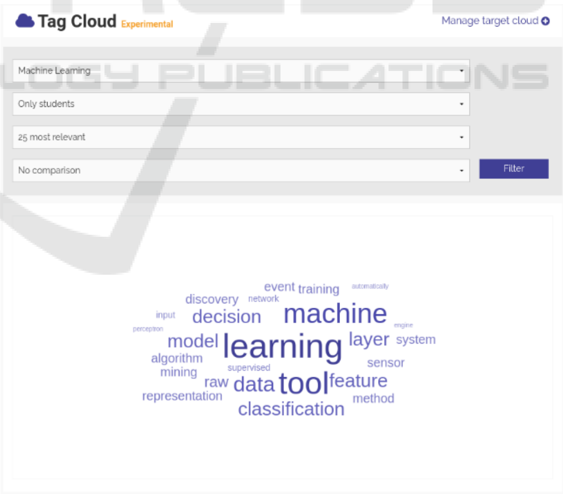

Figure 1: Word cloud of the "Machine Learning" topic.

As shown in Fig. 2, a teacher is able to create a target cloud per topic or course. They can type a new term and add it to the cloud or select one of the words they have already used during the topic or course and insert it into the cloud. The words used by the teacher are listed on the right column and sorted by frequency of usage. The teacher can also remove a word from the cloud at any time. In this case, we have built a target cloud with ten concepts related to machine learning: algorithm, classification, decision, knowledge, learning, mining, model, network, pattern and training. As it was mentioned before, if no weight is specified, it is set proportionally to the number of target words. Since the target cloud has 10 words, each word will weigh 10%.

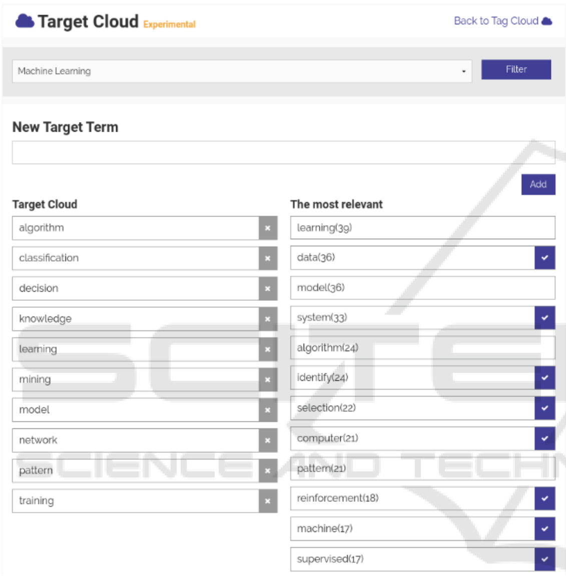

Figure 2: Target cloud customization view.

Once the target cloud is defined, the teacher can compare it with the topic word cloud. Back to the "Tag Cloud" view, we select the following options: Machine Learning topic, only students, 10 most relevant words and comparison with target cloud. It is important to remark that the number of words selected has a direct impact on the concordance calculation. To compare two clouds as accurately as possible, both must have the same size. That is why, in this case, only the 10 most relevant words from the topic word cloud have been selected.

Fig. 3 shows the mixture of word clouds generated after the application of the above filters. This mixture presents different colours for each word depending on the cloud they belong to: dark slate blue for those words belonging to the word cloud (i.e. Machine Learning word cloud), light grey for those words belonging to the target cloud and orange for those words belonging to both clouds. Just like the rest of the cloud representations, the size of each word is proportional to its frequency of use. Keeping all this in mind, let us analyse the concordance metric. Concordance is measured through words and their respective weights. The existence of each target word in the word cloud is validated and then the deviation between its weights is calculated. The smaller the absolute deviation of all weights, the better the concordance between the two clouds. The best case scenario is when students use all words from the target cloud at the same frequency/weight specified by the teacher. In the case in hand, the concordance between the two clouds is about 31.59%. There are only four words that both clouds have in common: learning, decision, model and classification; but their frequency of use is pretty similar to that expected from the teacher. That is the reason why, even though students use only a few target words, the concordance is acceptable.

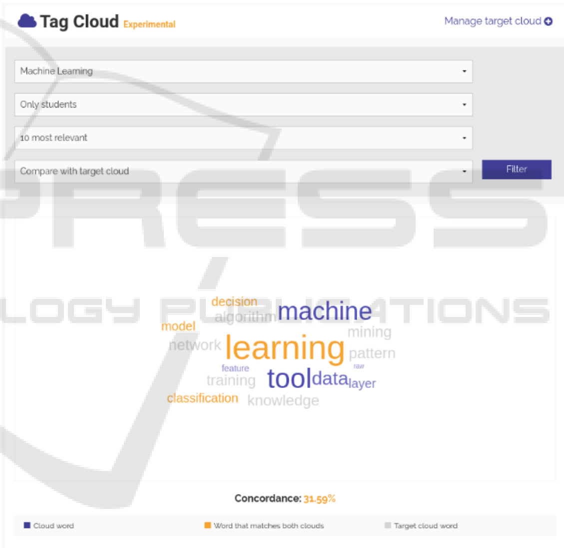

Figure 3: Comparison of "Machine Learning" word cloud with target cloud.

By modifying one of the previous filters, namely the user filter, we can get a comparison between a student word cloud and the target cloud. Fig. 4 and Fig. 5 shows two examples of these comparisons. The first one presents a peculiar case because: in this case there are six words belonging to both clouds, but not all of them have the expected frequency of use. That explains why the concordance here is only of 18.82%.

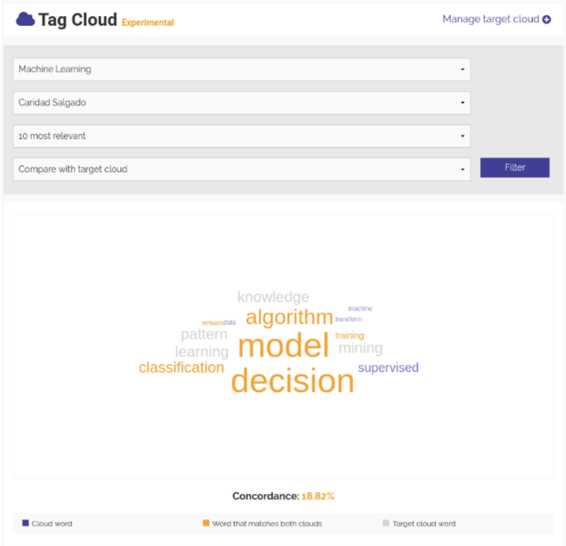

Figure 4: Comparison of “Machine Learning” word cloud with target cloud for a specific student.

Figure 5: Comparison of “Machine Learning” word cloud with target cloud for another student.

4 CONCLUSIONS AND FUTURE WORK

The Word cloud provides a visual and intuitive representation of knowledge from the student/s, allowing to create a fast description from their contribution in an online learning process. Education analytics could overload the teacher, but the word cloud makes it easier to understand what is going on in a course with a single eye span.

Word cloud can be used in gap analyses, showing what is missing and what is expected using the different colours to identify if the words are in the student cloud, the teacher cloud or both. With this information, the teacher could redirect the learning process introducing new materials and exercises to reinforce the misrepresented subjects in the word cloud.

In the future, we will expand these features also to be available for the students. It could be very interesting for them to know which concepts they have missed, but that the teacher expects to be well known. This could boost motivation to discover the ideas proposed by the teacher that they have left behind.

Also the distances from student’s word clouds to the teacher one could be useful to provide an automatic evaluation measure.

ACKNOWLEDGEMENTS

This research has been supported by Ministerio de Economía y Competitividad (Centro para el Desarrollo Tecnológico Industrial, Contract IDI-20161039), Junta de Extremadura (Contract AA-16-0017-1, and projects GR15106 and GR15011), Cátedra ASPgems, and European Union (European Regional Development Funds).

REFERENCES

Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., and Ananthanarayanan, V. (2017). *NMC Horizon Report: 2017 Higher Education Edition*. Austin, Texas: The New Media Consortium.

Aeiad, E., & Meziane, F. 2016. Validating the learning outcomes of an e-learning system using NLP. *Natural Language Processing and Information Systems*, *9612*, 292-300.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61-75). Springer New York.

Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching and Learning*, *8*(2), 25.

Calle-Alonso, F., Cuenca-Guevara, A., de la Mata Lara, D., Sánchez-Gómez, J. M., Vega-Rodríguez, M. A., & Sánchez, C. P. (2017). NeuroK: A Collaborative e-Learning Platform based on Pedagogical Principles from Neuroscience. In *Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017)-Volume* Vol. 1, pp. 550-555.

Czerkawski, B. 2016. Blending formal and informal learning networks for online learning. *The international review of research in open and distributed learning*, *17*(3), 138-156.

dos Santos, J. C. A., & Favero, E. L. 2015. Practical use of a latent semantic analysis (LSA) model for automatic evaluation of written answers. *Journal of the Brazilian Computer Society*, *21*(1), 21.

De Paolo, C. A., & Wilkinson, K. 2014. Get your head into the clouds: Using word clouds for analyzing qualitative assessment data. *TechTrends*, *58*(3), 38-44.

Feinberg, J. 2014. Wordle-Beautiful word clouds. *Wordle—Beautiful Word Clouds*. http://www.wordle.net/

DeNoyelles, A., & Reyes-Foster, B. (2015). Using Word Clouds in Online Discussions to Support Critical Thinking and Engagement. *Online Learning*, *19*(4), n4.

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014, January). Word cloud explorer: Text analytics based on word clouds. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 1833-1842). IEEE.

Kaptein, R., Hiemstra, D., & Kamps, J. (2010, January). How different are language models and word clouds. In *Advances in Information Retrieval: 32nd European Conference on IR Research (ECIR 2010)* (Vol. 5993, pp. 556-568).

McNaught, C., & Lam, P. 2010. Using Wordle as a supplementary research tool. *The qualitative report*, *15*(3), 630-643.

Miley, F., & Read, A. 2011. Using Word Clouds to Develop Proactive Learners. *Journal of the Scholarship of Teaching and Learning*, 11 (2), 91-110.

Nevado-Maestre, A., Vega-Rodriguez, M. A., Perez, C. J., Calle-Alonso, F., & Sanchez-Gomez, J. M. 2017 Social Network Analysis and Educational Data Mining for NeuroK e-Learning Platform. *Innovative and Creative Education and Technology*, 23-27.

Nickell, J. 2012. Word clouds in math classrooms. *Mathematics Teaching in the Middle School*, *17*(9), 564-566.

Perry, L. 2012. Using word clouds to teach about speaking style. *Communication Teacher*, *26*(4), 220-223.

Porter, M.F. 1997. An Algorithm for Suffix Stripping. *Readings in Information Retrieval*, 313-316.

Porter, M.F. 2001. *Snowball: A language for Stemming Algorithms.* http://www.snowball.tartarus.org/texts/introduction.html

Reeves, T. C. 2003. Storms clouds on the digital education horizon. *Journal of Computing in Higher Education*, 15(1), 3.

Yukselturk, E., Ozekes, S., & Türel, Y. K. (2014). Predicting dropout student: an application of data mining methods in an online education program. *European Journal of Open, Distance and E-learning*, *17*(1), 118-133