Applying machine learning for tag classification in a collaborative knowledge system

Aplicando o machine learning para a classificação de etiquetas em um sistema de conhecimento colaborativo

DOI: 10.55905/oelv21n8-140

Recebimento dos originais: 28/07/2023
Aceitação para publicação: 31/08/2023

Bruno Zolotareff dos Santos
Master in Communication
Institution: Escola Politécnica da Universidade de São Paulo (POLI - USP)
Address: Av. Prof. Luciano Gualberto, 380, Butantã – SP, CEP: 05508-010
E-mail: bzsantos@usp.br

Sandra dos Santos Vales
Doctor in Sciences
Institution: Escola de Engenharia de São Carlos da Universidade de São Paulo (EESC - USP)
Address: Av. Trabalhador São Carlense, 400, São Carlos - SP, CEP: 13566-590
E-mail: drinhavales@gmail.com

Jorge Rady de Almeida Junior
Doctor in Electrical Engineering
Institution: Escola Politécnica da Universidade de São Paulo (POLI - USP)
Address: Av. Prof. Luciano Gualberto, 380, Butantã – SP, CEP: 05508-010
E-mail: jorgerady@usp.br

ABSTRACT
The use of technological resources to enhance learning has grown exponentially in recent decades, mainly with the advent of mobile communication devices that increased the interactivity of users who collaborate to form a network of collective intelligence. In this technology environment that uses the Web, there is a large volume of data that is commonly disorganized, which is a challenge to use this data in the learning process in a continuous way to complete the knowledge. This study proposes using metrics capable of measuring knowledge aggregated in metadata shared in a collaborative system to be used in a recommendation system in the tagging process in continuous learning.

Keywords: knowledge, metrics, metadata, recommendation, learning.
RESUMO
O uso de recursos tecnológicos para melhorar a aprendizagem cresceu exponencialmente nas últimas décadas, principalmente com o advento de dispositivos de comunicação móveis que aumentaram a interatividade dos usuários que colaboram para formar uma rede de inteligência coletiva. Nesse ambiente de tecnologia que usa a Web, há um grande volume de dados que é comumente desorganizado, o que é um desafio usar esses dados no processo de aprendizado de forma contínua para completar o conhecimento. Este estudo propõe o uso de métricas capazes de medir o conhecimento agregado em metadados compartilhados em um sistema colaborativo a ser usado em um sistema de recomendação no processo de identificação em aprendizado contínuo.

Palavras-chave: conhecimento, métricas, metadados, recomendação, aprendizagem.

1 INTRODUCTION
Digital tools have become fundamental in the daily lives of learners and professionals involved in the field of education. Since the advent of Web 2.0 and its digital media resources, it has contributed to the advancement of the learning process in all areas of knowledge (CORDEIRO, 2020).

Therefore, there is no single way to assess and measure knowledge considering only one path to learning. Constructing ideas using different means of technology without considering what the learner already knows can be seen as a mistake in any learning process (AGRA, 2019; CONRAD; OPENO, 2019).

Some systems based on knowledge discovery are applied to large amounts of data on the Web and additionally used in social networks. The identified data, mainly in collaborative systems, such as social networks, undergo a dynamic process, changing the interpretation of metadata due to user interaction within a continuous process of collaborative learning.

It is necessary to use a new concept that considers not only the use of metadata to organize and retrieve information but also the quality of the metadata that is used by the student during the learning process. To establish which metadata to use considering the collective intelligence generated in a collaborative system, it is necessary to apply cognitive metrics capable of classifying the metadata at the level of knowledge.
The process of identifying and classifying metadata is one of the most critical factors to be considered in Human-Computer Interaction, as it involves the feelings and needs of users, making it necessary to consider the adversities in communication in the use of collaborative social systems (ELAHI et al., 2016).

The quest to discover knowledge and what to do and where to apply this knowledge is a challenge that involves a set of differentiated solutions and technologies, making knowledge of data the main value. Being a differential in the learning process the use of technologies to fill gaps in knowledge that the student needs to reach his goal, which is to master the knowledge of the object of study.

The evolution of analytical systems presents factors of great importance, in an era of the use of cognitive computing generated by collaborative means, in which value is added to knowledge based on human learning, being a powerful tool to assist in the learning process (LIMA, 2018).

The ability of systems to recognize the cognitive level in a collaborative network can leverage metadata because it involves a collective effort from many users in their interaction with metadata-free classification systems (VIANNA., 2019).

The large volume of data and the generation of knowledge in Web systems can be a great ally in the supervised learning process, however, for this data to be used correctly, Data Mining techniques are used in the summarization process, helping in the discovery of knowledge and mainly in the classification of data used in collaborative systems to help the continuous learning process (AHMED, 2019).

Part of the objective of the research is to provide a means of classifying metadata at the knowledge level to be used in a recommendation system as a way of offering metadata that represents the added value generated in a collaborative way, mainly in systems that use a tagging process, making it possible to organize learning with terms, which are keywords, facilitating the organization of ideas through the construction of a mental map using collective knowledge with the aid of technological tools.

Thus, it is intended to use cognitive metrics capable of measuring the knowledge generated collaboratively in a classification matrix, facilitating the understanding and the degree of importance of the set of metadata suggested by the recommendation system.
2 RELATED WORKS

The use of virtual systems for organizing and retrieving information and learning comes from principles of using metadata used in military projects that later were converted into software to search for scientific material used as a practical way of organizing studies related to the subject (WILSON, 2015). This advance in the use of markers for objects in online systems was due to principles and studies that were linked to some American scientists in the period after World War II. Theodore Nelson started studies on the interconnection between the documents, with the forerunner of hypertext. The scientist's project, known as Xanadu, was the first connection data retrieval system (XAVIER, 2011).

The practice of using tags on the web and sharing content became known as folksonomy, these features were used in specific social networks that served any user to use the tagging practice to organize their objects of study on the web, among them, Delicious stands out, which it was the first social network with this purpose that is no longer in operation. The use of online learning platforms has been prevalent in recent times, especially at the time of the pandemic about COVID-19, the relevance and effectiveness of the use of platforms and their aspects was studied to identify the main functionalities that of the learning system, among these studies, the MOODLE, Blackboard and Sakai platforms stand out. (KASIM, 2016; ZHONG, 2022).

The organization of content on learning platforms uses metadata standards that are the basis for structuring the data available on the platform, as can be seen in several studies (DAGGER, 2007; KASIM, 2016; TILLAEV, 2023). However, unstructured data within these platforms also use folksonomy as a subsystem, so that learners can freely use the tagging practice and circumvent the problem of finding and organizing data logically for each learner (PETER, 2011).

In addition to learning platforms and tagging to organize and retrieve information in the learning process, other strategies are used to organize metadata. a cognitive network being used in several areas of knowledge and applied in research (AL-JARF, 2022; TANTAM, 2006). The organization of metadata using a collaborative means of communication is not something new, since the implementation of online learning
systems and social networks the tagging resource is available most of the time, the problem is the lack of resources in the recommendation systems of these tags, are metadata that often do not represent knowledge and, when reused, cause engagement problems in the collaborative network, and as information related to metadata, some researchers use Data Mining techniques in systems to improve the quality of this metadata generated in the practice of tagging for different types of applications (CHELMIS; PRASANN, 2011; CORLEY, 2010; PIERCE, 2008).

The use of collaborative systems on the web with tagging subsystems is widely used, some data analysis is carried out with Data Mining techniques, and mind maps are used as a way of structuring knowledge in a cognitive network. However, the research proposal is to improve the quality of metadata offered in the recommendation system, in a way that the learner can identify which level of knowledge is visually added to this tag.

3 DETAIL PROBLEM STATEMENT

The process of identifying the knowledge of the meaning of the data is one of the most relevant factors, which is justified by several reasons. Among them are: the large amount of unstructured data in Web systems such as virtual social networks and the absence of cognitive metrics based on human knowledge in metadata classification systems.

Monitoring the student's communication process is one of the most relevant factors, this occurs for several reasons, including the large number of students in the classroom, and the lack of tools to support development monitoring for a correct adaptation in the virtual environment of study.

The student in the classroom or in a virtual teaching environment has difficulties in following the development of the content to reach the intended objective because they are not adapted according to the needs of each learner's profile, and the established sequence is sometimes not in accordance with according to the student's cognitive capacity, leaving gaps in the knowledge necessary for the construction of knowledge.

The environment and the way in which the content is exposed act as an inhibiting factor for some users of the learning environment and are often discouraging. The
performance evaluation must be continuous in the learning process, which is one of the important factors for the individual (BRANSFORD, 2000).

The importance of considering self-knowledge in a study process is very important for the use of collaborative learning tools, mainly because it makes available a large amount of unstructured data. The lack of tagging resources and moderate adaptation are factors that hinder continuous meaningful learning, and the few content recommendation resources are not related to the reality of the participating group, disregarding the collective knowledge that is generated in the learning process.

The interpretation of the use of metadata in Web systems is often confusing, however, the definition for its use in collaborative systems that use the markup practice has some variations. In our research, metadata is identified with terms that describe the meaning of use for tagging purposes and has visual identifiers that symbolize the level of collaborative knowledge that is aggregated in the choice tag, facilitating the process of indexing the object in posts, reducing errors from improper use of tags with uncommon and little-known terms or with a meaningless semantic formation that does not represent some generated knowledge.

4 KDD WITH CKL ALGORITHM

In order to use the collective intelligence data generated by the tagging system in a collaborative network such as a social network like the Twitter, it is necessary to use supervised Data Mining techniques to summarize the unstructured data available in public posts that represent the knowledge referring to a given study object. The CKL algorithm is proposed in this study for Data Mining and has the following steps considering the Knowledge-Discovery in Databases (KDD) illustrated in Figure 1:

1. **Selection:** After extracting metadata from tweets that are posts from the Twitter social network, separation is performed to be used in a sample.
2. **Pre-processing:** The sample is loaded into an instance in the database for analysis, some filters are applied to remove noise and the metadata is organized in a Bag of Words (BOW).
3. **Transformation**: Some filtering and categorization rules are applied in this phase the metadata is counted to verify the frequency, known as Term Frequency - Inverse Document Frequency (TF-IDF).

4. **Data Mining**: With the metadata prepared with the transformation phase, the CKL algorithm classifies the metadata at the level of knowledge using the rules established in its cognitive metrics that apply the Pearson Coefficient and expressions responsible for balancing values.

5. **Evaluation**: With the metadata categorized, the CKL algorithm performs the evaluation obtained from the indices generated by the metrics contained in its logic and labels its metadata results according to their magnitude, applying Jacob Cohen's evaluation rules; this practice is known as Entity Recognition Named Entity Recognition (NER).

Figure 1 - Steps by Knowledge Discovery in Databases.

For the application of KDD, it is necessary to consider several steps of the CKL algorithm that involve cognitive metrics in the classification process, the operation of the algorithm and the description of the metrics are detailed in the next sections.
4.1 METRICS FOR RANKING EXTRACTED METADATA FROM THE RECOMMENDER SYSTEM

The method used in this work to classify the metadata is based on the linearity metrics of Pearson's Correlation and Jacob Cohen's interpretation of magnitude in the ranges of values used as a reference for the results from -1.0 to 1.0, which is the margin used for this research (MCHUGH, 2012).

The range from -1.0 to 1.0 used in this research is based on the margins used in the Pearson Correlation calculations and its interpretation of values that belong to this scale is used as a reference of magnitude by Jacob Cohen, who establishes labels to interpret the level at which fits the data referring to the interval, being: ‘high’, ‘medium’ and ‘low’. Jacob Cohen's interpretation of correlation results as a way of interpreting magnitude with intervals that are (CORRELL, 2019; KRAFT, 2019; MCHUGH, 2012): values ≤ 0.0: indicating no agreement; 0.0 – 0.20: like none a level; 0.21 – 0.59: moderate; 0.60 – 0.90: strong.

The values established in the intervals and the labels to identify the status of the metric's meaning were defined by Jacob Cohen, which is used in this research and adapted according to the rule established for the metadata classification levels, adaptation carried out by other researchers who use calculations of correlation (KOTTNER, 2011; VANBELLE, 2016).

4.1.1. Interpretation of established metrics

The metadata suggested by the system has an added value to them, on a measurement scale generated by the knowledge metrics described below. The metrics developed in this article are based on KMA (Knowledge Monitoring Accuracy), KMB (Knowledge Monitoring Bias) and NAC (Level of Knowledge Acquisition) metrics (ARDIANTO, 2020; BLANCO-FERNÁNDEZ, 2011; SINGH, 2023). The three new proposed knowledge metrics are:

- CKM (Collective Knowledge Monitoring): measures the level of user agreement on the domain of use of the selected metadata, compared to the collective knowledge generated in the system by other users.
• **KMD (Knowledge Monitoring Deviation):** measures and identifies possible deviations in the monitoring of user knowledge in relation to the domain in the use of the selected metadata. It makes a prediction in relation to the input metadata chosen by the user, which can be agreement or disagreement, when compared to the collective knowledge generated in the collaborative system.

• **CKL (Collective Knowledge Level):** this metric indicates the degree of collective knowledge added to the metadata selected by the user in the search process.

### 4.1.2. Parameters of the applied metrics

The values used in the parameters are extracted from the frequency of the terms used in the metadata, which are shared in the collaborative system. These values obtained from the tagging process represent the collective knowledge generated in the system and which is the basis of the values used in the expressions established in this research.

To calculate the MCC and DMC metrics, it was necessary to create some attributes based on the Pearson Correlation scale that vary between an interval of -1.0 and 1.0. These are QF, A, B, C, D, and QG, where each variable serves only as an index, values assigned according to the rule established by the recommendation algorithm. Below is the meaning of the parameters used to obtain the values of the MCC and DMC metrics in Table 1:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning of the parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>QF</td>
<td>Frequency of the chosen metadata term in the system.</td>
</tr>
<tr>
<td>A</td>
<td>Verification if the chosen metadata term is among the first three positions of the vector</td>
</tr>
<tr>
<td>B</td>
<td>If the chosen metadata term belongs to the first position of the vector.</td>
</tr>
<tr>
<td>C</td>
<td>Whether the chosen metadata term is between the second and third positions.</td>
</tr>
<tr>
<td>D</td>
<td>If the chosen metadata term is among the first three positions of the vector.</td>
</tr>
<tr>
<td>QG</td>
<td>Frequency of sorted metadata set.</td>
</tr>
</tbody>
</table>

Source: The authors

To apply the developed methodology, two expressions (1) and (2) are used to find the CKM and the KMD. Since the CKM expression evaluates the level of correctness with the sum of the CKM and KMD expressions, it can vary from -1.0 to +1.0. In order
to arrive at a correlated value within the group classified by the algorithm proposed in this work, it is necessary to apply the Pearson Correlation calculations, which were adapted and are being represented by ($\Sigma xy$), values that represent the frequency of the metadata terms, which is compared by the term chosen by the user where its frequency and term is represented by ($\Sigma y^2$).

The values used to calculate the $M$ expression are obtained by the frequency of interactions represented by $(fn)$, all values generated collaboratively in the recommendation system, which are used in the correlation expression (KRAFT, 2020):

$$M = \frac{\Sigma xy}{fn\Sigma y^2}$$  \hspace{1cm} (1)

With the value obtained from $M$, it is possible to obtain an estimate of correlation, however, it is necessary to consider some aspects because it is a collaborative system. In order to mitigate any interference in the metadata classification process, it is necessary to use an expression of weight and measures to mitigate interferences that occur in the communication process, mainly because it is a collaborative environment that can cause interference in the tagging process. To mitigate any interference in the metadata qualification process, the following calculation of the expression is performed:

$$CKM = \frac{(M * (QF + [(A + B) /2 -0,50] + [(D + C) * -1,00])) / QG}{QG}$$  \hspace{1cm} (2)

$$KMD = \frac{(A * 0,50) + (B * -0,50) + (D *1,00) + (C * -1,00)) /QG}{QG}$$  \hspace{1cm} (3)

In order to apply these metrics within a system such as e-Folks, it was necessary to create a generic algorithm illustrated in Figure 2, capable of performing these functions, from ordering applying folksonomy, and suggesting tags at the level of knowledge.
The values assigned to variables A, B, C and D depend on the comparison between the metadata chosen by the learner with the metadata of the first four positions generated by the folksonomy of the system represented in Table 2.

| Table 2 - Parameters used in expressions CKM and KMD. |
|-----------------|-----------------|------------------|------------------|
| Vetor           | A               | B               | C               | D               |
| Position (1)    | 1,00            | 1,00            | 1,00            | 0,00            |
| Position (2)    | 0,50            | 0,50            | 0,50            | 0,00            |
| Position (3)    | 0,25            | 0,25            | 0,25            | 0,00            |
| Position (4)    | -1,00           | -1,00           | -1,00           | 0,00            |

Source: The authors

The sum of the CKM and KMD values is intended to regulate the estimate and deviation that the user has during the act of choosing the answer, in this case, metadata. This sum results in the CKL, which is the main metric that represents the knowledge level of the metadata according to the proposal.

\[
CKL = CKM + KMD
\]  

The value of CKL is related to the knowledge added to the metadata chosen by the user, and possible non-agreement with the metadata most indicated by the collective knowledge, considering the comparison with the metadata best positioned by the frequency. With the CKL value is based on the interpretation of Jacob Cohen interpreted in Table 3 (KRAFT, 2019):

<table>
<thead>
<tr>
<th>Table 3 - Interpretation of CKL values</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKL scale</td>
</tr>
<tr>
<td>Less than 0.29</td>
</tr>
<tr>
<td>0.29 to &lt; 0.49</td>
</tr>
<tr>
<td>Greater than 0.50</td>
</tr>
</tbody>
</table>

Source: The authors

These metrics are applied in the classification algorithm for the recommendation system proposed in this work.
4.2 DESCRIPTION OF THE PROPOSED ALGORITHM

The flowchart contained in Figure 2 represents the proposed algorithm, with its phases described below:

- **Start**: Metadata is generated by students in the collaborative platform directed to the object of study.
- **Tag Choice**: The learner chooses the search term that represents the object of study that he wants to index on the Web.
- **Folksonomy**: A comparison is made between the learner's knowledge and the collective knowledge generated in the learning recommendation system.
- **Suggestion of Tags**: After checking and applying the metrics, the system suggests three metadata at the knowledge level to be indexed in the form of tags.
- **Mental Map**: When the learner receives tag suggestions from the system, a mental map is generated that represents the metadata used in the tagging process in the learning period, which is continuous.
- **Evaluation**: The mind map can be evaluated by an expert advisor who can suggest some metadata that complete your knowledge.
- **End**: Once the metadata is added to the mind map, the learner completes the process and can update each interaction with the system.

![Figure 2 - Flowchart of the CKL algorithm.](image)

Source: The authors

4.3. METADATA CLASSIFICATION PSEUDOCODE

To carry out the steps described in the algorithm, it is necessary to manipulate the metadata and order each interaction according to the frequency of the term chosen by the system user. Below is the pseudocode that performs the tasks of classification, suggestion and application in the mind map, see in Table 3:
Table 3 - Pseudocode of metadata sorted and chosen.

Pseudocode that performs the tasks of classification

procedure ApplicationRecommendSystem()
begin
character folks[] = bubbleSort({a1,a2,a3,...,an}] // sorted metadata
character tag = “term” // entered by user
character ncc = [-1.0 to 1.0] // value established in the CKM and KMD variables
real cog[] = bubbleSort({a1,a2,a3,...,an}] // sorted metadata
real mentalmap[] = {a1,a2,a3,...,an}] // chosen metadata
if (cog[ncc] != 0) then
for (int i = 0; i < cog.size; i++) then
if (tag : cog[i<3]) then
printf(cog.size); // prints the metadata hint
if(cog[i] == true) then
int x = 1//escolhe de 1-3
mentalmap[] = cog[x]; // adds the metadata in the Mind Map in a vector
function bubbleSort(A : list of sortable items)
do
exchanged:= false
for each i in 0 to length(A ) - 2 do // check if they are in the right order
if A[i] > A[i - 1] then // swap elements from descending place
exchanged (A[i], A[i - 1])
exchanged:= true
end
Source: The authors

5 METHODOLOGY

For the application of the study, a set of metadata was selected and extracted from tweets (SANTOS, 2023), a filtering and separation process was applied to remove noise, and the metadata was organized in a Bag of Words (BOW), which were loaded in an instance in the vector in a database. This metadata underwent a supervised Data Mining analysis, with the comparison class being known. This study related to the popular understanding of what the object called COVID-19 is about, with the intention of using collective knowledge as a tool to complete knowledge during a continuous learning process.

The set of metadata was used as the basis of the memory generated by Collective Intelligence. To compare the user's knowledge during the selection process with the collective knowledge, making it possible to classify the search metadata into levels of knowledge.

These levels of knowledge are attributed to going through a long process of Data Mining using metrics capable of measuring the knowledge added to the metadata. The
output of the results of the proposed CKL algorithm is a set of metadata at the knowledge level. To apply these concepts, a prototype was developed as a collaborative study tool, which is presented in the next section.

5.1 PROTOTYPE FOR APPLYING THE METHODOLOGY

The prototype developed called e-Folks applies the practice of tagging using the metadata recommendation system in the form of tags so that the learner can use it during the process of studying the object.

To start the collaborative study, the advisor discloses the object of study to the learners who belong to the same study group, which is illustrated in Figure 3.

Figure 3 - e-Folks screen for learners to post their items.

The first post is made by the advisor, in the case of the example it is as a user called Steve, who in his post is putting the title and objective of the object of study. Since the title to be studied is: "Market Analysis - the study of cell phone sales" and then a subtitle is added with the question: "Which brands are most in demand and why?" adding the initial search term so that the recommendation system suggests coming from an instance with metadata extracted from places that contain information related to consumption and marketing that are social networks.

After the first post in the collaborative study system, the other apprentices using the system start to contribute important data in the learning process, adding content to
their posts and mainly indexing tags related to the subject. In Figure 4, it is possible to see that there are two more positions besides the advisor contributing to continuous learning collaboratively.

Figure 4 - e-Folks screen with posts already shared in the system.

The interaction of the learners with the collaborative learning platform generates metadata that are the tags indexed in the posts, enabling the construction of a mental map for each learner, where their tags will be used as a way of organizing the studied content and the tags of the other learners who can help build knowledge.

This prototype was used with an anonymous group of students to verify the proposed learning methodology using metadata at the knowledge level in a continuous
learning process, the results of the experiment are in the experiment application section. For the process of choosing tags, the e-Folks tool uses its metadata recommendation system illustrated in Figure 5.

Figure 5 - Tag recommendation screen at knowledge level.

![Tag recommendation screen at knowledge level](image)

Source: The authors

5.2 EXPERIMENTAL SETUP

- Experiment: The object of study was passed on to the learners with the Covid-19 theme, the same was suggested for the other participants, however, in this case, the study was oriented so that the participants reached the objective of understanding what it was about until have the understanding and mastery of the studied object. Making it clear that your study would be evaluated by the study advisor, in this case, one of the authors of the article who is a professor at this higher education institution. This experiment lasted thirty days, during which the two apprentices interacted with the e-Folks platform.

In the learners' interaction with the collaborative learning platform, some metadata were created and others reused suggested by the collaborative system. It is possible to see in the mental map illustrated in Figure 6 that learners 1 and 2 used a lot of metadata suggested by the collaborative recommendation system. While each learner generated their metadata that is shared across the system.
By analyzing the result of the metadata generated by the learners compared to the collaborative metadata suggested by the system, it can be considered that there is progress in studies during the learning process. The research does not take into account links and other objects added to the posts, in this case, the focus of the study is the use of metadata as a means of organizing studies and retrieving information.

The objective is for the learner to be able to fill in the knowledge gaps that remain to be completed, using mental triggers with the help of metadata, not needing to study content, just completing the knowledge to achieve the objective that is the purpose of meaningful learning (AGRA, 2019).

Achieving the objectives in the study process using a collaborative tool can enhance learning, improving understanding within a context directed to the object of study, as was done in the experiment with metadata.

It is possible to note from the analysis that only five metadata used by the learners were not suggested by the recommendation system, reaching a satisfactory result for an initial experiment with the e-Folks tool that applies the metrics suggested in this work and classifies a group of metadata that are suggested in the recommendation system.

The metadata used in the instance that is loaded in a vector was generated by the users and some terms were inserted by the advisors as a way of directing the learning of the algorithm used for the study related to the term Covid-19. Even though it was an experiment carried out in thirty days and used only two apprentices to investigate the
experiment, the work was valid because it was real data generated by users of the e-Folks system. The metadata used by the two learners and the metadata suggested by the recommendation system is illustrated in Table 4.

<table>
<thead>
<tr>
<th>Collaborative</th>
<th>Learner 1</th>
<th>Learner 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>accident</td>
<td>accident</td>
<td>accident</td>
</tr>
<tr>
<td>astrażeneca</td>
<td>arcturus</td>
<td>astrażeneca</td>
</tr>
<tr>
<td>china</td>
<td>china</td>
<td>astrażeneca</td>
</tr>
<tr>
<td>coronavirus</td>
<td>coronavirus</td>
<td>pandemic</td>
</tr>
<tr>
<td>infection</td>
<td>infection</td>
<td>'Johnson &amp; Johnson</td>
</tr>
<tr>
<td>laboratory</td>
<td>laboratory</td>
<td>pandemic</td>
</tr>
<tr>
<td>Pfizer</td>
<td>omicron</td>
<td>Pfizer</td>
</tr>
<tr>
<td>SARS-CoV-2</td>
<td>pandemic</td>
<td>SARS-CoV-2</td>
</tr>
<tr>
<td>search</td>
<td>Pfizer</td>
<td>Qingdao</td>
</tr>
<tr>
<td>vaccine</td>
<td>SARS-CoV-2</td>
<td>vaccine</td>
</tr>
<tr>
<td>Wuhan</td>
<td>search</td>
<td>Wuhan</td>
</tr>
<tr>
<td></td>
<td>vaccine</td>
<td>Wuhan</td>
</tr>
</tbody>
</table>

Source: The authors

Even being an experience with two apprentices in the learning process in a short period, it is necessary to consider that it is a model of supervised learning; not being necessary to have many apprentices to verify the functioning of the system, it can be considered an adequate performance directed to the objective that it is learning using technological resources with the interaction of other users in a collaborative way, using metadata classified in the level of knowledge that is a characteristic not found in other tools used in the learning process. The comparison results of the learners with the metadata suggested by the recommendation system show a similarity between the linearity of the chosen metadata, this result generated by the recommendation system.

6 CONCLUSION

The study of collaborative learning systems is significant for the construction of knowledge that with the help of collaborative tools can bring an excellent resource. Tools such as e-Folks created for this purpose explore little-explored fields, which is the use of collective intelligence in favor of the meaningful learning process in social networks with
a focus on learning. Despite being a commented subject on the importance of collective intelligence and the application of artificial intelligence tools to improve the capabilities of the tools, there are not many studies focusing on the quality related to metadata in collaborative learning systems, which, by analysis, are keywords responsible for the cognitive activation of understanding related to the object of study, being the semantic representation and human knowledge through natural language, which with the application of Data Mining techniques can summarize data in a specific way, improving the selection of metadata in the case of the e-Folks recommendation system. The study brought significant data for research, not yet fully conclusive regarding the effectiveness of using collaborative systems using the tag system at the level of knowledge, however, the experiment with learners using the e-Folks tool brings perspectives on the use of the CKL algorithm with their cognitive metrics, which generated enough metadata for a comparative analysis between users and the recommender system.

**FUTURE WORKS**

The e-Folks tool was created for the application of the CKL algorithm in a knowledge discovery process using Data Mining techniques, this tool can be improved and used in a permanent server and applied to a larger group of learners.
REFERENCES


KASIM, Nurul Nadirah Mohd; KHALID, Fariza. Choosing the right learning management system (LMS) for the higher education institution context: A systematic review. *International Journal of Emerging Technologies in Learning*, v. 11, n. 6, 2016.

KOTTMER, Jan et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *International journal of nursing studies*, v. 48, n. 6, p. 661-671, 2011.


