Knowledge Level Tags: Applied to Collaborative Recommender Systems on the Web

B.Z. Santos [a,*], J.R. Almeida Jr [b] and S.S. Vales [b]

[a] *Polytechnic School (Poli), University of São Paulo, Av. Prof. Luciano Gualberto 380, Butantã – SP, CEP 05508-010, Brazil*
E-mails: bzsantos@usp.br, jorgerady@usp.br

[b] *São Carlos School of Engineering, University of São Paulo, Av. Trabalhador São Carlense 400, São Carlos - SP, CEP 13566-590, Brazil*
E-mail: drinhavales@gmail.com

Abstract. This article aims to present a tag recommendation model at the knowledge level in a collaborative system on the Web. One of the main reasons for this proposal is due to limitations in the tagging process, causing loss in the quality of the terms used in the metadata that are indexed in posts on social networks in the form of tags, losing the meaning of the relationship between the tags and the object, resulting in a lack of engagement in the collaborative system by not exploring the potential of collective intelligence in a more practical and visual way to be identified by the user when choosing tags in the process of indexing the object. In this study, an algorithm for classifying metadata at the knowledge level is proposed, which uses metrics capable of measuring the collective intelligence aggregated to the metadata generated in the system, with two main steps being assigned, which are the classification and recommendation of a set of tags at the knowledge level.

Keywords: Knowledge, Semantic, metadata, metrics, social network

1. Introduction

Twitter, a social network known as microblogging, has grown at an accelerated since its launch in 2006. It is not just a communication environment with a social scope restricted only to a group of friends; it has become an innovative tool for sharing digital content, advertising, real-time news, and offering services [1]. Over time, Twitter has become one of the most widespread social networks, generating over 500 million daily tweets [2]. The informational content of Twitter has attracted many researchers over the years to study different social phenomena, trying to establish a correlation between the vast public opinion expressed on Twitter and physical events that occur in society. These research efforts encompass diverse areas, including product sales predictions [3] and the spread of infectious disease outbreaks [4], among other surveys. These efforts are usually divided into steps: filtering to classify tweets and then using supervised learning in the comparison step to finding similarities between tweets in the prediction process [5]. In the prediction stage, some researchers use regression algorithms to find the correlation between the reference content and the group of generated tweets to arrive at a result close to the expected one.

*Corresponding author. E-mail: bzsantos@usp.br.

One of the main challenges for researchers in the filtering process is to identify irrelevant tweets in large volumes for disposal since relevant tweets are scarce [6–8]. An approach used by many researchers involves organizing the terms used in tweets by adopting a classifier with textual characteristics that restrict the group of words or use only keywords depending on the analysis [9–12]. These classifiers are known as Bag-of-Word (BOW), which have been successfully employed in predictions with limited data in several types of analysis, such as user sentiment analysis performed with classifiers such as Naïve Bayes and Support Vector Machine [13, 14].

In some dynamic recommendation systems, metrics applied to collaborative filters are used in the knowledge discovery process [15, 16], which can improve the classification results of metadata by applying real-time values, as is the case with Twitter [2]. However, most recommendation systems do not focus on qualifying the level of knowledge aggregated in the metadata, which is primarily responsible for the engagement of tweets in the social network in the use of hashtags [17, 18], thus generating failures of interpretation and use in the tagging process. The use of metrics to measure the level of knowledge about metadata for content classification is an underexplored factor in recommendation systems [19, 20].

It is necessary to employ appropriate methods that enable the application of metrics in collaborative filters capable of quantifying the level of knowledge aggregated in metadata generated in the tagging process. Twitter is one of the social networks with limited resources in tagging practices; the same applies to other social networks, such as Facebook and YouTube, which lack visual resources to identify the level of knowledge contained in the tags, thus increasing the errors of choice by the system user when indexing some object with tags in posts. The main contributions of this article are: (a) propose a method of classifying and recommending metadata; (b) propose adopting a set of metrics to the knowledge level; (c) apply visual resources that are easy to identify to improve the interpretation in the tagging process; (c) use a prototype tool to test the proposed method.

2. Related works

Researchers have adapted some techniques to classify metadata in a web environment [17]. Wang et al. [21] used Wikipedia data as a repository in their analysis to identify and classify metadata relating to most of the titles on featured pages on the web. Kahn et .al [22] collected 3,955 metadata in tweets directly from the Twitter corpus, applied the filtering process, and reduced it to 887, not using other repositories to avoid the risk of adaptation problems. A similar technique had applied by Culotta [23], who selected four terms for specific keywords to find some relationship in a set of 206 tweets to classify into 160 positive examples and 46 negative ones to predict the intensity of the flu in New York, which served as the basis for using a Naive Bayes classifier with more tweets. Some researchers employed many fast filtering techniques to reduce the number of irrelevant tweets in the Twitter corpus to avoid noise, which are the irregularities found; these filters used terms as keywords and other filters on information contained in URL [9–11, 23]. In our experiment, the filtering process uses a similar technique to reduce irregular metadata and uses terms that are the keywords needed to make predictions according to the references.

There are other techniques used in the process of filtering collected metadata with a different approach using metrics. Ochoa et al. [24] created metrics to measure the learning level of users while using teaching tools that employ the practice of tagging. Király [25], based on metrics by Ochoa et al. [24] and Bruce et al. [26], devised a metadata quality assurance framework that applies some metrics to classify digital metadata to classify metadata considering the use of established standards such as Dublin Core (DC). The use of metrics, according to research, is recommended in the classification process of metadata collected in digital systems such as Twitter [24–26]. Elahi et al. [15] developed a collaborative filtering system focused on learning for recommendation aided by a learning algorithm, considering the ranking of a term's frequency in social networks. Collaborative recommender systems such as social networks can classify metadata by applying different techniques to find the best metadata correlation, using collaborative filters and conceptual models, which require adjustments for each nature of the application [16]. Our model uses collaborative filtering techniques aided by metrics capable of measuring the correlation of collective intelligence between collaboratively generated metadata, using an algorithm to assist in the automatic classification process attributed to the metadata recommendation system, applying various techniques presented from related research.

3. Definition of metadata applied to tagging practices in recommender systems

In our research, we use metadata generated in social networks that are known as keywords, which are terms that describe some object on the web, so the word 'term' serves to identify the name of the metadata, which according to some researchers, can be described as bookmark, tag, or hashtag [27, 28], all three are synonyms referring to the description of metadata used in the practice of tagging in social networks.

4. Detail Problem Statement

As mentioned in the related work section, using hashtags in tweets includes some of the filtering, ranking, and recommendation challenges discussed in this section.

4.1. Restrictions

Twitter users have adapted to the use of hashtags in their tweets that are shared on the social network, believing that they will be understood by the context and indexed content, estimating that the target audience is already familiar with hashtags. The only resource available to the human reader to understand hashtags in tweets is textual interpretation according to their level of knowledge. The use of hashtags has no restrictions, is free to use, and may cause ambiguity related to the context used. Twitter does not provide support regarding the meaning of hashtags with context, it only recommends the hashtags most used in tweets by other social network users, and it is up to the user to choose or not the hashtags recommended by Twitter. The use of hashtags serves as a way of organizing and retrieving unstructured content made available by users and finding a pattern in this type of practice is challenging. Hashtags used in tweets can generate dimensions necessary for the use of a Bag-of-Word (BOW) classifier; however, classifying hashtags by higher frequency of use will not always guarantee an appropriate choice to be used in tweets, as they do not have any rules emphasizing that they belong to the context. Tweets need to be more consistent in the use of hashtags, which may cause difficulties in content classification and understanding for other users of the social network who need to use hashtags in their tweets. For example, let us consider the specific hashtag 'apple' in three tweets:

1. "I buy a smartphone #apple"
2. "I love berries #apple #health"
3. "I love the song called apple of love #apple #fikar #musicao"

In the first tweet, the user is referring to the company that manufactures smartphones that has the name of Apple. In the second tweet, he refers to the apple fruit. In the third tweet, he refers to a song called apple of love. There are four cases related to these three tweets to mention:

(i) Tweets with different contexts (1), (2), and (3) use hashtags in common; regardless of the context, the same hashtag is used in a single instance for training.

(ii) The use of more than one hashtag in the same tweet (2) and (3), there is a relationship of contextual human interpretation. However, there may or may not be a direct association between the hashtags, but only one instance for training should be considered.

(iii) Tweets often have textual similarity, which could be considered for a better understanding of the use of hashtags by users. Nonetheless, there are unusual contexts of the use of the same hashtag as in tweets (1), (2), and (3) that are considered in the same training instance.

(iv) In tweets, it is common to appear hashtags whose meaning has no direct connection with the context, sometimes with a formatting error or whose purpose is little known by most users (3); a single training instance is used even with ambiguity regarding hashtag-related errors.

In case (i), the choice of a common hashtag is associated with the collective intelligence that the hashtag 'apple' refers to a company, and which is used by many users in their tweets as it is part of a collaborative system where within a context makes sense to read the content. In case (ii), the choice of using more than one hashtag, 'apple' and 'health', can improve the understanding of the context in which it is being used. The hashtags can have a human interpretation association in relation to the text they refer to. However, the hashtags used may not have a typical interpretation relationship. In case (III), hashtags may or may not have some textual similarity in common with other tweets, and unique hashtags may be combined that may hinder human interpretation in relation to the association of the context used by most users of the system. In case (iv), some hashtags like 'musicao' and 'fikar' do not have a direct connection with the hashtag ' apple '; they have to do with the user's feelings in relation to the text, causing some inconsistency in relation to the writing, or they are hashtags with little aggregated collective intelligence, making the classification process difficult. Therefore, in case (iv), in the tweet example (3) refers to confusing hashtags.

Definition 1 Confusing Hashtags: The use of confusing hashtags in tweets can happen in three situations:

a. When preparing the text for the tweet, the user needs more knowledge about which hashtag best represents the indexing of the content that will be shared on the social network. b. The user is influenced by Twitter's hashtag recommendation system, where the set of hashtags is offered when the user starts the hashtag indexing process in the tweet. c. The user is simply unaware of the usefulness of hashtags as a collaborative tool, and the user is not intended to reach a target audience or for any commercial marketing purpose; hashtags are constructed with incoherence, errors, and sometimes phrases.

It is possible to decrease the error rate of a classification system by removing outliers related to the tweet (3) identified as confusing hashtags [29]. In this article, we propose a method with some procedures for selecting the necessary features to locate and adjust values used in the training set. The results obtained in the evaluation show that only a few outliers are generated with the hashtags in the tweets that can impact the proposed method. The discrepancies are eliminated in the training set to improve classification accuracy [30].

4.2. Leakage of engagement

Twitter is a tool where people publish news, opinions, and thoughts with some domain. The mastery of some issues varies over time, especially concerning understanding social network users; this can happen due to changes in physical events that can approximate or distance the relationship between events [29]. The use of hashtags is related to the domain of the subject, and its understanding can also vary according to time and may or may not make sense during a period [24]. For example, in the evolution of computers in the 80s and 90s, the term used to describe the technological device was microcomputer, which over time came to be called just computer, both referring to the personal computer [31]. Therefore, examples of tweets: "We used to use computers to program in Basic in the 80s: https://youtu.be/seM9SqTsRG4 *#microcomputer* or "The personal computer is being replaced by smartphones: https://theuktime.com/smartphones-replace-laptops-and-pcs/.

The two tweets do not have problems with the use of hashtags. However, the problem is related to the domain of using hashtags, which impacts the engagement of tweets on the social network, not reaching the target audience [16–18]. In the first tweet, the hashtag:'*#microcomputer*' is probably little known by users on the social network. In a search for hashtags in the Twitter tool, we found that the recommendation system does not suggest any hashtag as a recommendation; since the hashtag: '#computer', when inserted in the tweet, the recommendation system offers three more related hashtags, which are: '*#Computertechnology*', '*#computerart*' and '*#computers*'. The Twitter tool probably did not recommend other hashtags about '*#microcomputer*' because it did not achieve enough repetitions in other tweets, which is common in ranking methods [30]. The Twitter tool uses only the most used and shared tweets as a ranking factor, and its recommendation system only suggests hashtags that have the suffix or prefix in the hashtag initially inserted in the tweet [32, 33].

To try to solve the problem of engagement with the hashtags used in tweets and to increase the possibility of reaching their target audience, many users add more than one hashtag so that the tweet can better associate the collective intelligence of the social network. For example, if in the first tweet where the hashtag '#microcomputer' was used, two new hashtags were added, such as '#computer' and '#PC', it would probably increase the understanding of more users of the social network by reaching audiences from different eras: "We used to use computers to program in Basic in the 80's: https://youtu.be/seM9SqTsRG4 $\#microcomputer\#computer\#PC$. Many users are unfamiliar with the hashtag '#microcomputer' because it refers to something older [31]. When using more than one hashtag in the same tweet, the user does not receive suggestions from the Twitter recommendation system for hashtags related to the first hashtag. Therefore, hashtags are chosen randomly, hoping to reach the target audience without considering any relationship between the hashtags referring to the collective intelligence, which directly impacts tweet engagement on the social network [6]. In the proposed method, the hashtags used in the tweets are related and categorized by applying classification metrics regarding the input term chosen by the user, being selected by a set of metadata from the same training instance, which according to studies of coherence of semantic analysis can bring better similarity results between terms [24].

4.3. Lack of metrics for qualifying tweets

Hashtags are terms indexed in tweets that describe other data; in other words, they are metadata defined by their use and definition used in the semantic web [34, 35]. The concern about the quality of the terms used in the metadata has become something difficult to be observed by specialists, as it needs to be a taxonomy. The classification of objects in the posts freely has become a problem because it is something difficult to control, even if it presents benefits of organizing and retrieving information on the web, it still needs attention to be more effective in the choice of terms in the use of metadata [13, 14]. The standard criterion used as a metric for suggesting metadata on social networks is the frequency of the term used in posts such as tweets. However, using free-mode metadata on the web generates some errors: poor formation of the term used in hashtags, lack of identification of use in context, and mistakes in the correlation between the term and the indexed object [36]. There needs to be metrics to identify the best metadata option suggested by the recommendation system of social networks to help choose the term of use on the web; no index identifies which would be the best option for engagement on the social network. For example, the hashtags $\#java$, $\#javacoffe$, and #javascript in the tweet are static: "The most popular computer programming language: https://youtu.be/Tck7MFXRQZg $\#java\#javacoffe\#javascript$. The social network's recommendation system could have suggested these hashtags at the time of the search, as it is not possible to visually verify which of the hashtags represent the best metadata with their terms, making something difficult to choose or identify, in addition to the indexed content that may be outdated or containing some context error like the hashtag $\#javacoffe$ [24]. One of the objectives of this work is to classify the metadata with indexes generated by classification metrics that show your level of knowledge related to the social network.

5. Metrics for ranking extracted metadata from tweets

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 to reference the results from -1,0 to 1,0, the margin used for this research. These values mean the intensity of the correlation strength that can be positive, negative, or values close to zero, which means a weak correlation [37].

Jacob Cohen recommends the interpretation of the correlation results represented in an objective way as a way of interpreting the correlation magnitude of the coefficients with intervals that are: values $\leqslant 0,0$ as indicating no agreement, $0,0$ – $0,20$ as none to slight, $0,21$ – $0,39$ Minimum; $0,40$ – $0,59$ weak; $0,60$ – $0,79$ moderate; $0,80$ – $0,90$ strong and above $0,90$ almost perfect [38, 39]. These values established in the intervals and the labels to identify the status of the meaning of the metric defined by Jacob Cohen and are used as a basis for other studies with different application purposes, such as Schuster [40], Vandebelle [41], Kottner [42], Warrens [43] and in recent studies by Grandini [44], Chicco [45], Al-garadi [46]. These studies propose adapting labels and margins to interpret the correlation magnitude for practical reasons in the studies involved [41].

5.1. Application of the proposed metrics

The proposed metrics aim to improve the quality of metadata used in posts on social networks, offering the user a set of qualified metadata at the level of knowledge where the term used in the metadata has a better understanding of the collective understanding, helping in the content rating process. The metrics used make up part of the responsible recommendation algorithm. The metrics proposed in this section are used in a recommendation algorithm that generates a set of metadata at the level of knowledge offered to the social network user while choosing hashtags; the metadata terms are submitted and ordered according to their level. Despite using a large amount of data in your instance of a set of metadata $|C|$ in this research, one of the criteria established for filtering is the frequency of terms ***(ft)***, which are categorized from highest to lowest represented by $\mathbf{a_i}$.

$$Conjunct(metadata) = \frac{\sum_{i=1}^{n} a.ft_{(i)}}{\sum_{i=1}^{n}} \tag{1}$$

Where:

• ***(ft(i)*** is 1 if the i-ith field has a non-null value, 0 otherwise.

• N is the number of instances in which the term is present. To be able to classify the metadata generated in the Twitter social network, it was necessary to create some metrics suggested for this work:

KLE - Knowledge Level Estimate: measures the level between the agreement of the specific term used in the metadata chosen by the user, compared with the collective intelligence generated in the collaborative system by other users.

KLA - Knowledge Level Adaptation: measures and identifies possible deviations in monitoring the user's knowledge about the domain using the selected term.

MKL - Metadata Knowledge Level: This metric indicates the degree of collective intelligence added to the metadata the user selects in the search process. Its result is the sum of the two other metrics, KLE and KLA.

5.2. Parameters used in the suggested metrics

To calculate the KLE and KLA metrics, it was necessary to create some parameters: QS, PCV, IAV, AVA, IAV and RQ, which vary between an interval of -1,0 and 1,0 based on the Pearson Correlation scale. Values are assigned according to the rule established by the recommendation algorithm used in this research. The values are generated according to the comparison between the term (t) used as input by the user compared to the first four positions of the metadata set $|C|$ in the vector $||.||$, being analyzed in a single instance by default established in the research that is represented by N.

The meaning of the parameters used to obtain the values of the KLE and KLA metrics follows:

• *QS – Quantity Shared:* frequency of terms *(ft)*, chosen by the user used in the metadata, the term *(t)* needs to exist in the metadata set $|C|$.

$$QS = \sum_{i=1}^{n} ft_{(ti)} \tag{2}$$

• *PCV – Precision Correlation Value:* the term *(t)*, chosen by the user, does not belong to the metadata set $|C|$ in the vector $||.||$. That is, the system did not identify any relationship with the collective Intelligence generated by the system.

$$PCV = \sum_{i=1}^{n} t_i \notin ||.||_1 \leqslant 4 \tag{3}$$

• *IAV - Intermediate Accuracy Value:* the term *(t)* chosen by the user has some relationship between the set of metadata $|C|$ in the vector $||.||$. It presents an average level of knowledge correlated with collective intelligence.

$$IAV = \sum_{i=1}^{n} t_i \in ||ft_t||_1 > 1 \;\; and \;\; ||ft_t||_1 \leqslant 4 \tag{4}$$

• *AVA – Adequacy Value for Accuracy:* the term *(t)* chosen by the user has a relation to the set of metadata $|C|$ in the vector $||.||$. It presents a high level of knowledge correlated with collective intelligence, that is, the system identified a precise relationship with the typed metadata.

$$AVA = \sum_{i=1}^{n} t_i \in ||ft_t||_1 \geqslant 1 \;\; and \;\; ||ft_t||_1 \leqslant 4 \tag{5}$$

• *IAV – Intermediate Adequacy Value:* the term *(t)* chosen by the user has a relation to the set of metadata $|C|$ in the vector $||.||$. The metadata entered by the user presents an average level of knowledge correlated with collective intelligence, that is, the system identified an approximate relationship with the metadata entered.

$$IAV = \sum_{i=1}^{n} t_i \in ||ft_t||_1 > 2 \;\; and \;\; ||ft_t||_1 \leqslant 4 \tag{6}$$

• *RQ - Relative Quantity:* frequency of terms *(ft)* which belongs to the first position of the set of metadata in the vector $||.||$.

$$RQ = \sum_{i=1}^{n} t_i \in ||.||_1 < 2 \tag{7}$$

The values established for the parameters used to calculate the metrics vary according to the comparison of the term chosen by the user at the time of analysis with the first four terms of the metadata set $|C|$ in a single instance, considering the position *(P)* of the metadata, obeys the following scale of values established in Table 1:

The QS and RQ parameters do not follow the same scale, and their values are obtained during the filtering process. To assign the QS value, the recommendation algorithm suggested in this work compares the frequency of the term chosen by the user as an initial reference with the set of metadata already filtered and classified according to the frequency of the terms within the analyzed instance. As for the RQ parameter, the assigned value is chosen the term

Table 1
List of Subjects and grades obtained

Positions	IAV	IAV	AVA	PCV
P[1]	1,00	1,00	1,00	0,00
P[2]	0,50	0,50	0,50	0,00
P[3]	0,25	0,25	0,25	0,00
P[4]	-1,00	-1,00	-1,00	0,00

that has the highest frequency within the set of classified metadata. If there is not enough metadata in the database to perform the classification, the RQ value cannot be obtained, and consequently, the KLE and KLA values cannot be obtained. The values used in the metrics in their parameters to classify the metadata have a reference interval similar to that used in the interpretation of the magnitude suggested by Jacob Cohen and other researchers, which assigns known qualitative values such as the description of the magnitude of the ranges of values that uses descriptive names to identify the level of knowledge according to ranges between -1,00 and 1,00 [38]. For this research, the content of values represents a scale referring to the collective intelligence added to the set of metadata created by users of Twitter, related to the frequency of the term of metadata shared and reused in the social network to index the content in the form of hashtags, the following values assigned according to the intervals established for this search: the value 1.0 has a is considered to be satisfactory, that is, it represents the best value of collective intelligence. A value of 0.5 is considered to be adequate, meaning a slightly lower stage than the best value of collective intelligence. The value 0,25 is considered moderate, and 0,25 represents the lowest value of collective intelligence. On the other hand, a value with a low correlation between -1,0 means little or no related collective intelligence. Therefore, according to the position of comparison of the input term chosen by the user with the first four terms of the set of metadata classified by the recommendation algorithm, the results are assigned to the parameters at the time of analysis of a group of metadata classified according to the frequency of the term.

The intention is not to study how the user uses his knowledge to define the metadata but to measure the degree of relationship between the user's knowledge and the collective intelligence generated in the semantic network to help classify and suggest metadata. The KLE measures the hit estimate, and the KLA measures some user hit deviation. The sum of the two metrics results in the MKL, which is the value obtained in this knowledge scale determined in the model.

5.3. Correlation of values used in metrics

The KLE calculation takes place in part by obtaining a correlation between the values of the metadata set represented by $(\sum xy)$, in which they are compared with the reference value of the user's chosen metadata term $(\sum y^2)$. The values used to calculate the R expression are obtained by the frequency of metadata terms and interactions represented by (n), generated collaboratively on the social network with tweets; these values generated by the recommendation algorithm suggested in this research are used by the variables in the correlation expression adapted to the application context:

$$R = \frac{\sum xy}{n \sum y^2} \tag{8}$$

With the value obtained from R, it is possible to get a correlation estimate; however, it is necessary to consider some aspects because it is a collaborative system, and the higher frequency of terms may indicate a lower effectiveness of communication, presenting some type of noise that could divert or interfere with the understanding of the term when sharing the metadata on the social network, disturbing the knowledge in the process of choosing the metadata, this happens because there are factors that contribute to the appearance of noise in the communication, one of them is the adverse environment, where there is an excessive movement of elements that compromise attention, in this case during the process of indexing objects in tweets, where there is interference from much information at the same time, which is sometimes difficult to interpret [47, 48]. To mitigate any interference in the metadata qualification process, considering the effectiveness of the communication based on parameters established in the metric and the correlation values, the following calculation of the expression is performed:

$$KLE = (R * (QS + [(IAV + IAV)/2 * -0,50] + [(PCV + AVA) * -1,00])/RQ) \tag{9}$$

The KLE metric expression multiplies IAV and IAV values by -0.50 because they are considered intermediate precision values. It means that they have some relationship between the set of metadata but need a higher level of knowledge in relation to collective intelligence. Therefore, they are multiplied by a negative weight to indicate that their contribution to communication effectiveness is less than that of higher precision values such as the AVA and that their contribution negatively affects communication effectiveness.

The AVA and PCV parameters are multiplied by -1.00 because they represent the adequacy ratio of the terms the user chooses with collective intelligence. When AVA is more significant, it means that the selected term has a relationship with a better degree of accuracy with collective intelligence. At the same time, PCV indicates no correlation or relationship between the term and collective intelligence. Multiplying AVA and PCV by -1 means that these variables will negatively impact the final result of the expression. These variables indicate a need for more relationships or an imprecise relationship between the term chosen by the user and the metadata set. Thus, when multiplying them by -1, the formula inverts the value of these variables, indicating that a smaller value for these variables will positively impact the result of the expression.

This makes sense, as a lower value for AVA and PCV means a weaker relationship or no relationship, which suggests a lower communication effectiveness, the QS and RQ values consecutively represent the frequency of the metadata chosen by the user and the highest frequency of the metadata term among the four of the metadata set, obtaining a direct correlation without considering the other metadata.

This condition refers to the comparison process of the metadata term chosen by the user as a reference that will compare with the position from one to four of the vector, which, according to the position where the similar metadata is found, will generate the values that will be assigned to the parameters IAV, IAV, AVA and PCV, where part of the calculation obtains a result with adjustments taking into account the levels of accuracy compared to collective intelligence directly without using other frequency values of the additional metadata to know how much adjustment is necessary.

According to the values assigned in the parameters where the result can be highly positive or negative, being multiplied by the Pearson R value to obtain a result corresponding to the correlation of the other metadata that belong to the same set, already considering an estimate of success in the process of choosing the reference metadata with the additional metadata of the vector, where some uncertainty factors in relation to the measurement are taken into account , which was suitable for the proposed values of the parameters, the calculation of the KLA is given by the following expression:

$$KLA = ((IAV * 0,50) + (IAV * -0,50) + (PCV * 1,00) + (AVA * -1,00))/RQ \tag{10}$$

The KLA metric has the same parameters as the KLE metric; however, the objective is to find the deviation of knowledge related to the answer, identifying the user's characteristics when using his knowledge. The KLA expression uses intermediate accuracy and adequacy values (IAV and IAV) and accuracy and adequacy correlation values (PCV and AVA) to calculate the level of knowledge, weighting the IAV and IAV values positively and the PCV and AVA values negatively. KLA is a metric that measures the hit deviation compared to the metadata term chosen by the user with the others in the metadata set; with these values obtained in the KLE and KLA metrics, it is possible to calculate the MKL with the following expression:

$$MKL = KLE + KLA \quad (11)$$

With the MKL value that can vary within a scale between -1 and 1, values that are used in Pearson's correlation calculations and used as a basis for the scale in this research, which establishes intervals that can be interpreted in a relational scale with the correlation magnitude of the coefficients of the results to classify the metadata in a level of knowledge in a qualitative way like MKL: low, medium and high. Based on Jacob Cohen's interpretation interpreted in Table 2 [37]:

Table 2
Interpretation of MKL values

MKL scale	**Classification**	**Interpretation**
Less than 0,25	low MKL	The metadata term chosen as a reference has a low level of knowledge compared to the collective intelligence.
$0,25$ to $< 0,50$	average MKL	The metadata term chosen as a reference has an average level of knowledge compared to the collective intelligence.
Greater than 0,50	high MKL	The metadata term chosen as a reference has a high level of knowledge compared to the collective intelligence.

The value of MKL is related to the knowledge added to the metadata chosen by the user, and it is possible not to agree with the metadata most indicated by the collective intelligence. These metrics are applied in the classification algorithm for the recommendation system proposed in this work.

5.4. *Description of the proposed algorithm*

The developed algorithm uses the metrics proposed in this work to measure the level of collective intelligence aggregated to the metadata about the user's knowledge. However, the method can accept other metrics during the knowledge generation process, being flexible. The flowchart contained in Figure 1 represents the proposed algorithm, with its phases described below:

• **Start:** The system user is an agent who interacts to start choosing the term for the search metadata used as a reference; your decision to choose is free.

• **Fetch Metadata:** The system user can choose to use the new metadata he generated or select some metadata based on the already classified system's three suggestions if metadata is available in the database for that generation.

• **Insert in the Database:** After choosing the metadata by the user, the metadata is inserted in the database. This database is always updated in each interaction with the system; metadata is always inserted, whether new (entered by the user) or a value suggested by the system.

• **Applies Folksonomy:** The metadata stored in the database must be sorted in descending order, according to frequency. In this process, a set of inference rules is applied so that the metadata is related to all groups to which they belong relative to collective intelligence.

• **Calculates Knowledge Metrics:** calculates the measure and evaluates the level of knowledge added to the metadata and later qualifies if the metadata has a satisfactory index. The calculated metrics are described in the previous section: KLE (Knowledge Level Estimation), KLA (Knowledge Level Adaptation), and MKL (Metadata Knowledge Level).

• **Stored in the Database:** The metadata evaluated by the MKL is stored.

• **Assigns the Level Satisfactory:** The metadata has a high level of knowledge.

• **Assigns the Level Adequate:** The metadata has a medium level of knowledge.

• **Assigns the Level Moderate:** The metadata used has a low level of knowledge.

• Generates the Image of the Marker: According to the MKL, a tag (Satisfactory, Adequate or Moderate) is attached to be displayed with the hashtag.

• **Display Metadata Suggestion:** A set of classified metadata is suggested to the user with an identification marker of the level of aggregated knowledge.

• **Select Metadata:** The user can choose one or more metadata to use in the content indexing process in a post on the social network among the options suggested by the recommendation system.

• **End:** Ends the process of choosing the metadata; the user cannot select another

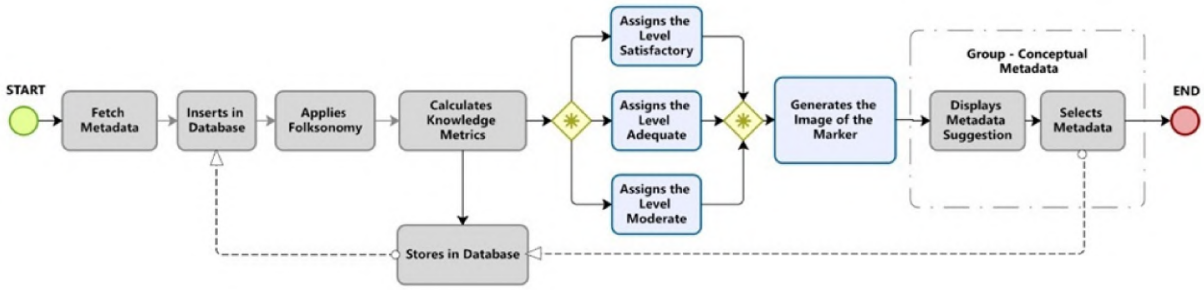

Fig. 1. Generic flowchart of the metadata classification algorithm.

5.5. *Applying indexes with MKL*

The attribution process of indexes that represent levels of knowledge is selected according to the MKL, a metric developed to classify metadata and is described in the previous section. To clarify the nomenclature used in the Algorithm 1, where it reads $'|C|'$ represents the first four metadata with the highest frequency of the term $'ft'$ in the vector, represented by $'X1'$.

The categorization of terms considering the degree of importance $'ai'$ is related to the frequency of terms 'ft(t)' organized in disbelieving order, reference to the expression (1) that represents the set of metadata for analysis. The MKL metric is assigned to the variable $'atrib_metrics'$, and the indices: $'Satisfactory'$, $'Adequate'$, and $'Moderate'$ are given to the variable $'Ind'$, which is added to the metadata and recommended to the system user. The indices are assigned according to the MKL related to the $'Z'$ scale, which ranges from -1,0 to 1,0, represented in the following pseudocode.

Algorithm 1 MKL Indexes

$t \leftarrow new\_term$
$|C| \leftarrow expression\_numer(4)$
$Z \leftarrow scale(\breve{}1,0\ to\ 1,0)$
$Ind[] \leftarrow indexes : 0 - \text{`}Satisfactory', 1 - \text{`}Adequate', 2 - \text{`}Moderate'$
$X1 \leftarrow Conjunct(metadata)$
$: atrib\_metrics \leftarrow MKL$
if t == true **then**
  while $|C| \in X1$ **do**
    if $t \in |C|$ **then**
      $metadata \leftarrow atrib\_metrics$
      if $metadata > 0,50$ **then**
        output $metadata \leftarrow Ind[0]$
      else if $metadata > 0,25\ and < 0,50$ **then**
        output $metadata \leftarrow Ind[1]$
      else
        output $metadata \leftarrow Ind[2]$
      end if
    end if
  end while
end if

6. Corpora of tweets

The metadata used for our recommendation system was extracted from the Twitter social network using the Trackmyhashtag tool, which extracts hashtags directly from tweets [49]. A total of 743,000 metadata were extracted using the terms 'apple14', 'iphone12', 'iphone' and 'apple' as references in the search. Metadata were filtered to exclude noise that could interfere with the results of training and tests carried out in the research, being categorized according to the rules established in filtering, with a total of 105 thousand metadata being categorized. In the filtering process, part of the exclusion criteria was the poor formation of the words used in the terms, writing errors, languages other than English, compound words and initial numbers before writing the term. Twitter was chosen due to its importance in the practice of tagging tweets and its relevance in profile analysis and knowledge discovery, especially in the marketing area [50].

7. Application of the recommendation system using prototype

A prototype was developed to conduct analysis tests of the suggestion process of a set of metadata that uses markers to identify the level of knowledge. The prototype consists of a social network called Cognomy designed as a microblog; this tool was developed based on the social network Twitter as a way to simulate the proposed recommendation system aimed at tweets from the social network, in which users make posts with messages of up to two hundred characters, which uses the metadata recommendation system generated by the classification algorithm suggested in the search.

7.1. General functioning of the prototype and graphical interface

To use the Cognomy prototype, a set of metadata already categorized by the subject was inserted into the database so that the information system could execute the classification processes established by the proposed method. The sequence of user interaction with the prototype is described below:

Cognomy: The user must register the information in the fields Title, Detail the subject and Type Tag as in Figure 2; after that, the system suggests the tags (according to the term chosen for the search tag entered by the user). The user types the Search Tag and clicks the Consult button to be redirected to the next Figure 3.

Fig. 2. Microblog screen for posting in the Cognomy prototype.

Tag suggestion: For simulation, a thousand metadata already categorized for this sample used in the experiment were added, data related to the same context with Apple products. This simulation is performed using the narrow folksonomy concept that limits the context in which the metadata is being used, considering only a group that generates metadata of the same subject [16, 51]. The term 'iphone' was chosen to be typed as a reference in the generated some bulleted tag suggestions. Suggested tags are: 'S iphone14', 'A apple' and 'M iphone', according to Figure 3:

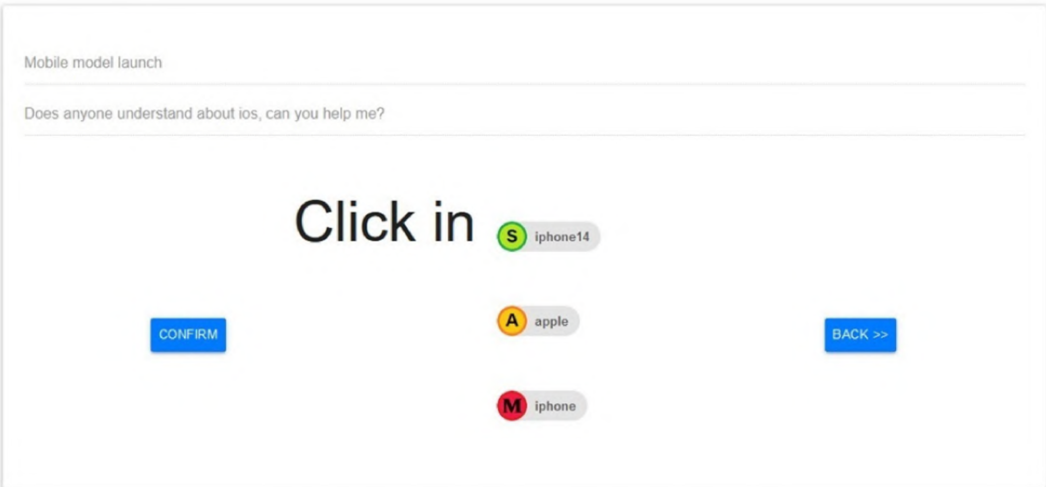

Fig. 3. Tag suggestion screen by the Cognomy system.

Postage: After choosing the tags that will be indexed Figure 4, the post is published on the Cognomy page, and each post is identified as having been generated by the user logged into the prototype. It should be noted that the post is public, as shown in Figure 5.

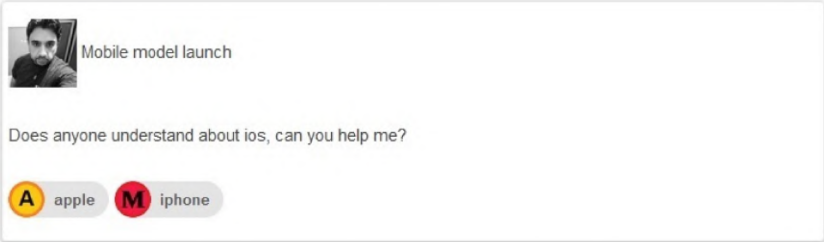

Fig. 4. Tag suggestion screen by the Cognomy system.

Upon completing the process of indexing the tags chosen to compose his post, the user has his message published on the Cognomy screen, where all publications generated by network users are found. The microblogging system with some posts can be seen in the illustration in Figure 5.

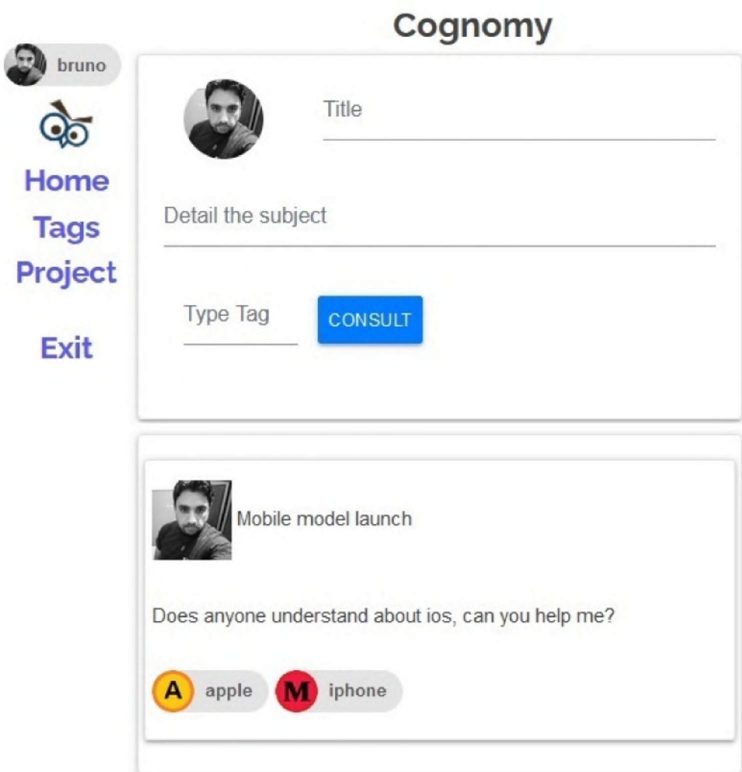

Fig. 5. Tag suggestion screen by the Cognomy system.

This tool was developed to test and verify the use of indexes embedded in posts in a microblog similar to Twitter. An experiment was carried out with a group of users to study the use of the tag recommendation system and verify its applicability; the results obtained can be found in the following sections of this article.

8. Results produced in the analysis stage for metadata classification

Twitter is a dynamic social network that receives a large amount of unstructured data in its tweets, which are constantly updated. Due to the frequency of use and sharing of the terms used in hashtags, it became necessary to investigate how the dynamics of the frequency of these terms work in relation to the amount of metadata related to the search term; for this experiment, a scalability curve was drawn.

8.1. Metadata scalability curve

A curve was drawn in relation to the amount of metadata related to the term 'iphone' to verify the scalability of the metadata within a period; this analysis has an interval every thousand metadata, totaling ten thousand metadata, analyzing the frequency of the term to investigate the entropy according to quantity [15], organized in Table 3.

Table 3

Amount of metadata in relation to term frequency

Quantity of metadata	Term frequency
1000	62
2000	65
3000	141
4000	221
5000	322
6000	362
7000	370
8000	375
9000	378
10000	381

According to the data presented in Table 3, the tendency is for the use of the chosen term to appear more often with the amount of metadata until it reaches stability with slight variation in values in relation to frequency. Applying the calculation of R2 with the result of almost 96% of scalability, it is possible to establish an application with samples of a set of metadata within a reliable range for sample preparation with the values shown in Figure 6.

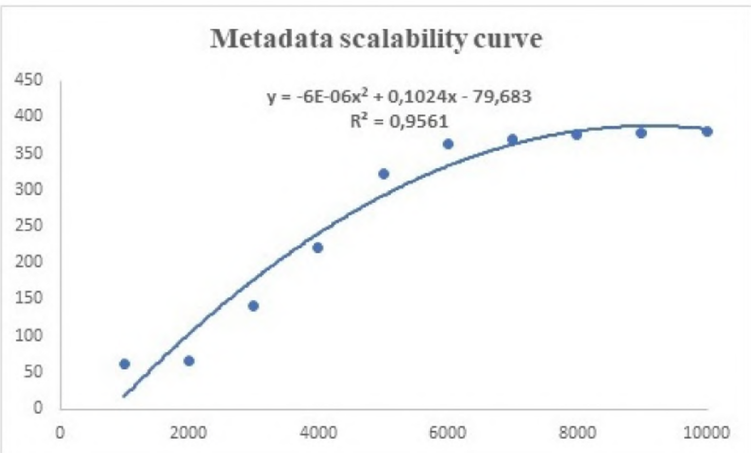

Fig. 6. Twitter metadata scalability curve.

8.2. Experimental setup

For this experiment, metadata extracted from tweets between July 2022 and April 2023 was used, similar to the approach by Culotta [23] and Aramaki et al. [52], who consider the use of the most relevant terms used in tweets in the set of corpora metadata used. A total of 105,000 metadata in the corpora were used, which are sufficient values according to the studies presented in relation to the observation of the frequency of terms used as reference. For these analyses, the following terms were used as references: 'iphone14', 'iphone12', 'iphone' and 'apple'. Five samples were separated in total, dividing each piece with eleven thousand metadata.

8.2.1. Classification of metadata by knowledge level

To arrive at the results of the MKL (Metadata Knowledge Level) five samples were used, dividing each sample with eleven thousand metadata, the mean values and the variance values were measured to investigate the deviation from the mean that the data of a set analyzed feature using the following expression.

$$S = \frac{\sum(x_i - \overline{x})}{n-1} \tag{12}$$

Table 4 below shows the result obtained when applying each sample to the proposed recommendation algorithm to find the MKL, the samples are numbered from 1 to 5 with the acronym of the MKL metric.

Table 4

Result of the metrics and variance obtained in the analysis of the MKL samples

Metadata	**MKL1**	**MKL2**	**MKL3**	**MKL4**	**MKL5**	**Average**	**σ (Variance)**
iphone14	0,30	0,27	0,25	0,29	0,27	0,28	0,044360698
iphone12	0,21	0,19	0,18	0,17	0,19	0,19	0,050739216
iphone	0,12	0,10	0,13	0,12	0,12	0,12	0,007784989
apple	0,29	0,26	0,27	0,28	0,24	0,27	0,042496708

It is possible to notice that the terms of the metadata have a small variance interval and that the best results are directed to the terms 'iphone14' and 'iphone12' because they are the most commented items by users on Twitter related to the launch of these products, the terms used are in the first positions for presenting a strong engagement in the social network in relation to the product and the brand during the period of this research. Among these results obtained from the MKL, only the three highest values represent the metadata with the highest level of collective intelligence that are recommended to the user in the tagging process that is part of the proposed recommendation system.

8.2.2. Experiment 1: Comparison of results with other algorithms

To compare the data obtained with the application of the recommendation algorithm used to obtain the MKL (Metadata Knowledge Level), two other algorithms were chosen, which are MultilayerPerceptron and SMOreg due to the following characteristics for analysis:

The **Multi Layer Perceptron (MLP)**, also known as Multilayer Perceptron (PMC), is a type of neural network composed of one or more hidden layers, containing a variable number of neurons in each layer. The gradient descent algorithm then updates the synaptic weights towards error minimization. In addition, neurons such as sigmoid activation functions also involve mathematical calculations to transform input signals into non-linear outputs.

SMOreg can handle extensive training sets and high dimensionality of the data, making it a popular choice for regression tasks where the data is complex and varied. SMOreg aims to find the hyperplane that minimizes the sum of squared forecast errors.

The mentioned algorithms were applied using the Weka data mining tool, which is available under a free license, mainly for academic use [53]. To establish an instance with the metadata set for analysis, a MySQL database connection was used and the same amount of metadata used to obtain the MKL results was loaded.

Table 5 illustrates the values obtained in the models applied for the study carried out in this work. The MultilayerPerceptron models, SMOreg compared to MKL showed similar results with little variation in correlation and linearity values, as illustrated in Figure 7. In Table 6, it is possible to verify the degree of variance that the recommended metadata values have in relation to the mean, which it is the main factor to be considered in relation to the collective intelligence that owns the selected metadata.

Table 5

Comparison of analysis results with MultPerc, SMOG and MKL

Metadata	MKL	MultPerc	SMOreg
iphone14	0,281817078	0,366729124	0,457400000
iphone12	0,235038753	0,054529788	0,495200000
iphone	0,194957115	0,320913964	0,197000000
apple	0,275858500	0,366372887	0,444000000

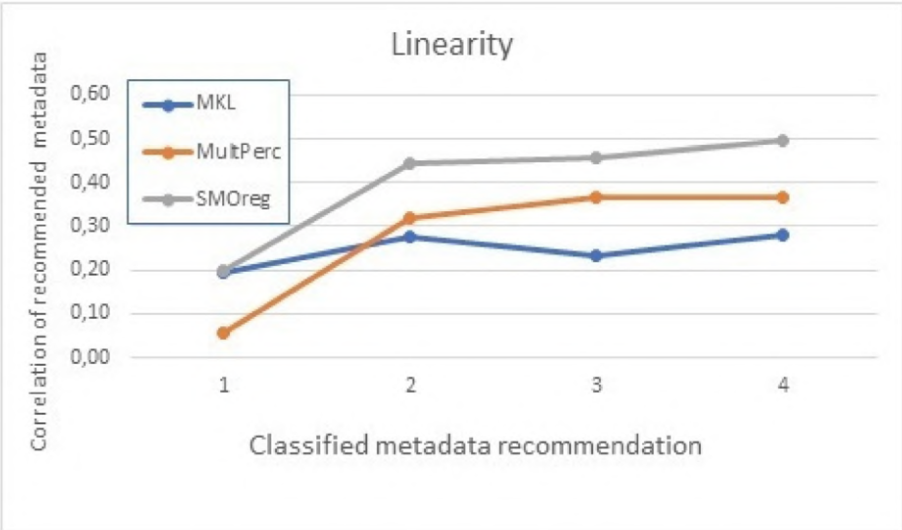

Fig. 7. Linearidade dos MKL, MultilayerPerceptron e SMOreg.

Table 6

Mean and variance of MKL, MultilayerPerceptron and SMOreg.

Models	σ (Variance)
MKL	0,0444
SMOreg	0,1228
MultPerc	0,1702

8.2.3. Results and discussion in experiment 1

The proposed model presents better linearity in relation to the metadata recommended by the algorithms; as illustrated in Figure 7, the correlation of the metadata in the same group presents close results between the models and the differences are in the adjustments made to find the MKL. The correlation values of the MultilayerPerceptron and SMOreg models' metadata are higher than MKL; however, they do not show linearity between the metadata. To be used as a suggestion, they may present discrepancies and not recommend the metadata properly; for example, in the MultilayerPerceptron, the correlation of the metadata with the term 'iphone12' is low, has a correlation (0,054529788), and the other metadata of the same group have a correction close to (0,35), a considerable difference, and that if it were applied to the proposed recommendation system, this value would probably discard the metadata. In SMOreg, there is a discrepancy in the correlation in the metadata with the term 'iphone'; it correlates (0,197000000), and the mean correlation of the other metadata is (0,45).

These deviations shown in MultilayerPerceptron and SMOreg compared to MKL may represent a lower strength of engagement of the metadata in the social network, losing linearity and not classifying the metadata appropriately; for this reason, the MKL presents better values in relation to the set of compared metadata to the other models. In Figure 7, the MultilayerPerceptron and SMOreg models do not demonstrate uniform linearity in relation to the straight line, even showing close correlation results. The three models present approximate values with little variation. The MKL x MultilayerPerceptron has a variation (σ: 0,1289), MKL x SMOreg (σ: 0,1285), and SMOreg x MultilayerPerceptron (σ: 0,1884). It is possible to verify the comparison of the correlation variance of the metadata in each model that are represented in Table 6, and the MKL presents the best value of the variance in relation to the mean of correlation, which means a better performance in relation to measuring the linear agreement between the terms used in tweets.

8.2.4. Experiment 2: Experiment with the Cognomy tool

For this experiment, the same set of metadata extracted from tweets between July 2022 and April 2023 reported in previous sections was used, with 105,000 metadata being loaded into the MySQL database instance to be used in the Cognomy microblog prototype. The tests were conducted anonymously with a group of two hundred volunteers among adolescents and adults aged between 15 and 45 years old.

This experiment was carried out to verify the effectiveness of the metadata recommendation system at the level of knowledge in the form of tags, which visually present knowledge indexes. The environment was prepared for posts aimed at Apple products focusing on the iPhone. There were two hundred and thirty posts in general on the Cognomy prototype, with more than half of the posts using more than one tag recommended by the system. Comparison to the set of metadata ordered by the system, according to its frequency represented by f(series). It is possible to visually verify in Figure 8 that the knowledge generated by users in relation to the use of tags is related to the collective intelligence extracted from the tweets.

Table 7
Results of the experiment with the Cognomy tool.

Metadata	f(serie1)	f(serie2)
iphone14	139	399
iphone12	97	267
apple	84	333
iphone	73	396

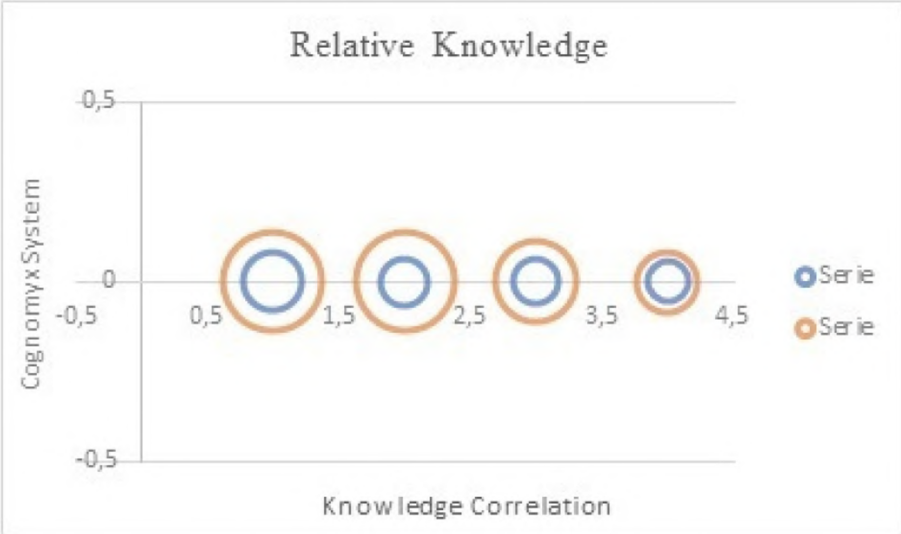

Fig. 8. Comparison of knowledge about a sample of tweets.

8.2.5. *Results and discussion in experiment 2*

In Table 7, the four most used terms of the metadata in the form of tags are the same ones the tool classified and recommended for the students. However, the only difference is in the frequency of use in the posts, where in this experiment, the term 'iphone' was more used than the term 'apple' compared to the classification of f(series). Even though not many users use the Cognomy tool in relation to Twitter users, it is necessary to consider that the application is being directed to a specific group with the same purpose, reducing interference with other subjects and external factors. For testing the use of the recommendation system proposed in this research, the implementation results were achieved with this experiment, recommending a set of metadata in the form of tags at the level of knowledge in a visual way to be easy to identify, reaching satisfactory values taking into account the goals set. Due to the number of tags used in the posts by the users, it is assumed that the recommendation made by the system helps in the tagging process, using a better engagement in relation to the collective understanding of the users of the social network. It is possible to notice a correlation between the tags chosen by the users and the tags recommended by the system. It is clear that this application is purposely targeted with the scope of verifying the use of the recommended tags in relation to Apple products. It is not the intention at this time of the research test in an open system, such as the test in the Cognomy tool, was carried out with a reasonable number of people within a period sufficient to reach the use of more than one hundred tags to verify that users are using the tags of the recommendation system.

9. Conclusion

This article argues that the hashtag recommendation system for tweets needs a knowledge-level classifier. Most systems require more resources in the recommendation process. Therefore, special measures must be taken when training a classifier for tweet hashtags. We propose two steps here; one, improve the selection of terms used in the metadata that is indexed as hashtags in tweets, removing noise and categorizing according to their frequency associated with the group referring to the term used as a reference, in which the user determines, classifying a group of metadata at the level of knowledge that has better linearity of its values in relation to the engagement of metadata in the social network. Therefore, a semi-supervised approach is adopted to create this set of metadata at the knowledge level, using the user's interaction with the system, taking into account aspects of communication in the use of metadata and relating the collective intelligence generated in the social network collaboratively, with the user's choice being recommended by the system, this set of metadata is sorted by the recommendation algorithm. We also proposed a metadata selection method not only for its frequency or similarity but for its correlation of metadata values identified by indices referring to the level of knowledge aggregated to the metadata, making this level of knowledge available visually in the tags, improving understanding and facilitating identification in the selection process.

References

[1] B.J. Naaman M and L. C, Is It Really About Me? Message Content in Social Awareness Streams, *CSCW '10 Proc. 2010 ACM Conf. Computer Supported Cooperative Work,* (2010), 189–192–, https://doi.org/10.1145/1718918.1718953.

[2] D.I. Gungor KN Ayhan EO, Tweet and account based spam detection on twitter. In: Artificial Intelligence and Applied Mathematics in Engineering Problems: Proc. Int Conf Artif Intell Appl Math in Eng (ICAIAME 2019), *Springer International Publishing* **ICAIAME**(10) (2019), 898–905, http://dx.doi.org/10.1007/978-3-030-36178-5_79.

[3] S. S, Evaluation of Tweet Sentiments Using NLP, *In: Int Symp Intell Inform: Proc of ISI 2022. Singapore: Springer Nature Singapore* **ICAIAME 2019** (2023), 225–238, https://doi.org/10.1007/978-981-19-8094-7_17.

[4] P.J. Azzaoui AE Singh SK, SNS big data analysis framework for COVID-19 outbreak prediction in smart healthy city, *Sustain Cities and Soc* **71:102993** (2021), 225–238, https://doi.org/10.1016/j.scs.2021.102993.

[5] P.D. Fonseca G O Hespanhol RM, Análise mercadológica do segmento pet: estudo de caso utilizando aprendizado de máquina, *Revista Latino-Americana de Inovação e Engenharia de Produção* **5(8)** (2017), 117–135, http://dx.doi.org/10.5380/relainep.v5i8.56495.

[6] F. S, Twibot-20: A comprehensive twitter bot detection benchmark, *Proc of the 30th ACM Int Conf Inf Knowl Manag* **30** (2021), 4485–4494, https://doi.org/10.1016/j.scs.2021.102993.

[7] A.Z. Alenezi MN, Machine learning in detecting COVID-19 misinformation on twitter, *Future Internet* **13(10)** (2021), 244, https://doi.org/10.3390/fi13100244.

[8] B. Ashutosh, Mining social media text for disaster resource management using a feature selection based on forest optimization, *Comput Ind Eng* **169** (2022), 108280, https://doi.org/10.1016/j.cie.2022.108280.

[9] W.X. Guo J, Image classification based on SURF and KNN, *IEEE/ACIS 18th Int Conf Comput Info Sci (ICIS). IEEE* **18th** (2019), 356–359, https://doi.org/10.1016/j.scs.2021.102993.

[10] M.T. Modha S Majumder P, An empirical evaluation of text representation schemes to filter the social media stream, *J Exp Theor Artif Intell* **34(3)** (2022), 499–525, https://doi.org/10.1080/0952813X.2021.1907792.

[11] P. AD, Classification Using Backpropagation Neural Network on Tweet Emotion Intensity, *Int Sem Appl for Technol of Info Comm (iSemantic). IEEE* **18th** (2021), 307–311, https://doi.org/10.1109/iSemantic52711.2021.9573240.

[12] D. S, Hybrid email spam detection model using artificial intelligence, *J Mach Learn and Comput* **10(2)** (2019), 316–322, https://doi.org/10.18178/ijmlc.2020.10.2.937.

[13] A. R, Sentiment analysis on E-sports for education curriculum using naive Bayes and support vector machine, *Jurnal Ilmu Komputer dan Informasi* **13(2)** (2019), 109–122, https://doi.org/10.21609/jiki.v13i2.885.

[14] S.B. Surya PP, Sentimental analysis using Naive Bayes classifier, *2019 International conference on vision towards emerging trends in communication and networking (ViTECoN) IEEE* (2019), 1–5, https://doi.org/10.1016/j.scs.2021.102993.

[15] R.N. Elahi M Ricci F, Image classification based on SURF and KNN, *IEEE/ACIS 18th Int Conf Comput Info Sci (ICIS). IEEE* **18th** (2019), 356–359, https://doi.org/10.1016/j.scs.2021.102993.

[16] S. T, Programming collective intelligence: building smart web 2.0 applications, *O'Reilly Media* **ISBN: 9780596550684** (2007), https://www.oreilly.com/library/view/programming-collective-intelligence/9780596529321/.

[17] W. V, Maximizing the tweet engagement rate in academia: analysis of the AJNR Twitter feed, *Ame J Neuroradiol* **38 (10)** (2017), 1866–1868, https://doi.org/10.3174/ajnr.A5283.

[18] D.L. F, Can social media improve stakeholder engagement and communication of Sustainable Development Goals? A cross-country analysis, *Technol Forecast Soc Change* **177** (2022), 121525, https://doi.org/10.1016/j.techfore.2022.121525.

[19] B.-F. Y, Exploring synergies between content-based filtering and spreading activation techniques in knowledge-based recommender systems, *Inf Sci* **181(21)** (2011), 823–4846, https://doi.org/10.1016/j.ins.2011.06.016.

[20] F. Z, Recommendation systems: Algorithms, challenges, metrics, and business opportunities, *Appl Sci* **10(21)** (2020), 7748, https://doi.org/10.3390/app10217748.

[21] L.X. Wang P, Assessing the quality of information on wikipedia: A deep-learning approach, *J Assoc Info Sci Technol* **71(1)** (2020), 16–28, https://doi.org/10.1002/asi.24210.

[22] S.K. Khan MAH Iwai M, An improved classification strategy for filtering relevant tweets using bag-of-word classifiers, *J Info Proc* **21(3)** (2013), 507–516, https://doi.org/10.2197/ipsjjip.21.507.

[23] A.C.S. Pierri F Artoni, Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections, *PloS one* **15(1)** (2019), e0227821, https://doi.org/10.1371/journal.pone.0227821.

[24] D.E. Ochoa X, Automatic evaluation of metadata quality in digital repositories. International journal on digital libraries, *IEEE/ACIS 18th Int Conf Comput Info Sci (ICIS). IEEE* **10** (2009), 67–91, https://doi.org/10.1007/s00799-009-0054-4.

[25] K. P, Towards an extensible measurement of metadata quality, *Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage* **2th** (2017), 111–115, https://doi.org/10.1145/3078081.3078109.

[26] H.D. Bruce TR, *The continuum of metadata quality*, ALA Editions, 2004, https://hdl.handle.net/1813/7895.

[27] B.J. Tajbakhsh MS, Semantic knowledge LDA with topic vector for recommending hashtags: Twitter use case, *Intelligent Data Analysis* **23(3)** (2019), 609–622, https://doi.org/10.3233/IDA-183998.

[28] B. B, Repertoires of remembering: A conceptual approach for studying memory practices in the digital ecosystem, *Journal of Communication Inquiry* **44(2)** (2020), 178–197, https://doi.org/10.1177/0196859919852080.

[29] S. Carter and A.C.-M. Yeo, Internet-enabled Collective Intelligence as a Precursor and Predictor of Consumer Behaviour, *Economics, Management, and Financial Markets* **13(4)** (2018), 11–38, https://doi.org/10.22381/EMFM13420181.

[30] K. U, Predictive analysis on Twitter: Techniques and applications, *Emerging research challenges and opportunities in computational social network analysis and mining. Springer, Cham* **18th** (2019), 67–104, https://link.springer.com/chapter/10.1007/978-3-319-94105-9/textunderscore4.

[31] R. Benjamin, A path forward for leveraging social media to improve the study of community resilience, *International Journal of Disaster Risk Reduction* **59** (2021), 102236, https://doi.org/10.1016/j.ijdrr.2021.102236.

[32] A. Ika, Utilizing hashtags for sentiment analysis of tweets in the political domain, *Proceedings of the 9th international conference on machine learning and computing* **9th** (2019), 43–47, https://doi.org/10.1145/3055635.3056631.

[33] K.M. Alghamdi MA, Intelligent analysis of Arabic tweets for detection of suspicious messages, *Arabian Journal for Science and Engineering* **45** (2020), 6021–6032, https://doi.org/10.1007/s13369-020-04447-0.

[34] G. R, Intermediary XML schemas: constraint, templating and interoperability in complex environments, *International Journal of Metadata, Semantics and Ontologies* **14(2)** (2020), 88–97, https://doi.org/10.1504/IJMSO.2020.108314.

[35] B. LA, Tagging: People-Powered Metadata for the Social Web (Smith, G.; 2008) [Book Review], *IEEE Transactions on Professional Communication* **18th** (2019), 321–322, https://doi.org/10.1109/TPC.2009.2025299.

[36] S.W. Peters I, Folksonomy and information retrieval, *Proceedings of the American Society for Information Science and Technology* **44(1)** (2019), 1–28, https://doi.org/10.1002/meet.1450440226.

[37] K. MA, Interpreting effect sizes of education interventions, *Educational Researcher* **49(4)** (2019), 241–253, https://doi.org/10.3102/0013189X20912798.

[38] M.M. L, Interrater reliability: the kappa statistic, *Biochemia medica* **22(3)** (2012), 276–282, https://doi.org/10.11613/BM.2012.031.

[39] C. J, Avoid Cohen's 'small','medium', and 'large'for power analysis, *Trends in cognitive sciences* **24(3):** (2019), 200–207, https://doi.org/10.1016/j.tics.2019.12.009.

[40] S. C, A note on the interpretation of weighted kappa and its relations to other rater agreement statistics for metric scales, *Educational and Psychological Measurement* **64(2)** (2019), 243–253, https://doi.org/10.1177/001316440326019.

[41] V. S, A new interpretation of the weighted kappa coefficients, *Psychometrika* **81(2)** (2016), 399–410, https://doi.org/10.1007/s11336-014-9439-4.

[42] K. J, Guidelines for reporting reliability and agreement studies (GRRAS) were proposed, *International journal of nursing studies* **48(6)** (2011), 661–671, https://doi.org/10.1016/j.ijnurstu.2011.01.016.

[43] W. MJ, Corrected Zegers-ten Berge coefficients are special cases of Cohen's weighted kappa, *Journal of Classification* **31** (2014), 179–193, https://doi.org/10.1007/s00357-014-9156-9.

[44] V.G. Grandini M Bagli E, Metrics for multi-class classification: an overview, *Cornell University* (2008), 05756, https://doi.org/10.48550/arXiv.2008.05756.

[45] J.G. Chicco D Warrens MJ, The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment, *IEEE* **9** (2021), 78368–78381, https://doi.org/10.1109/ACCESS.2021.3084050.

[46] A.-G. MA, Natural language model for automatic identification of intimate partner violence reports from twitter, *Array Elsevier* **15th** (2019), 100217, https://doi.org/10.1016/j.array.2022.100217.

[47] A.A. Kolajo T Daramola O, Real-time event detection in social media streams through semantic analysis of noisy terms, *Journal of Big Data* **9(1)** (2022), 1–36, https://doi.org/10.1186/s40537-022-00642-y.

[48] E.E. Taghandiki K, Implementation of a noisy hyperlink removal system: A semantic and relatedness approach, *Cornell University* **2303** (2019), 03321, https://doi.org/10.48550/arXiv.2303.03321.

[49] R.C. Truong H Salib A, The Use of Social Media in Pediatric Urology—Forging New Paths or Crossing Boundaries?, *Current urology reports* **20th** (2019), 1–8, https://doi.org/10.1007/s11934-019-0928-y.

[50] W.D. Naraine ML Wear HT, User engagement from within the Twitter community of professional sport organizations, *Managing Sport and Leisure* **24(5)** (2019), 275–293, https://doi.org/10.1080/23750472.2019.1630665.

[51] B. A, Lost in spotify: folksonomy and wayfinding functions in spotify's interface and companion apps, *Popular Communication* **18(1)** (2020), 1–17, https://doi.org/10.1080/15405702.2019.1701674.

[52] W.S. Aramaki E, Image and Sound of the City. In: The Social City: Space as Collaborative Media to Enhance the Value of the City, *Singapore: Springer Nature Singapore* **39** (2023), 205–214, https://doi.org/10.1007/978-981-19-7311-6_13.

[53] J.P. Dol SM, Use of Data mining Tools in Educational Data Mining, *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)* (2022), 380–387, https://doi.org/10.1109/CCiCT56684.2022.00075.