Secondary Nomination and Co-Referencing of Medical Terms in the Strategy of Harmonizing Indicators of Knowledge Assimilation in the Doctor's Portfolio

Ozar P. Mintser,[a] Larysa Babintseva,[a] Olga Sukhanova[a]

[a]Shupyk National Healthcare University of Ukraine

ABSTRACT

Some issues of the formation of the portfolio of doctors during continuous professional development are considered. Particular attention is paid to the peculiarities of the evaluation of the analyzed information, in particular, the problems of secondary nomination and co-referentiality of medical terms. The purpose of the study was to substantiate the prerequisites for creating a new type of portfolio by using methods of intellectual analysis of multidimensional information, as well as the formation of models and cognitive structures, which should become a central event in person-oriented active learning. It is shown that in order to ensure an adequate mechanism for evaluating multidimensional information in the doctors' portfolio, it is advisable to use ensembles of algorithms for intelligent analysis of big data. Conclusions: 1. The portfolio of doctors and pharmacists is of exceptional importance for the functioning of the system of continuous professional development of the relevant specialists. It helps organize lifelong learning and allows you to keep all evidence of learning and professional activity. 2. Taking into account the importance of the portfolio, it should be capable of standardized and formally calculated evaluation. 3. The entry of information into the portfolio is preceded by the determination of the semantic equivalence of the available information regarding the acquisition and assimilation of knowledge. 4. It is proposed to use ensembles of approaches to the analysis and arrangement of information in the portfolio - a combination of several algorithms that function simultaneously and provide an opportunity to correct possible errors. 5. The proposed decision-making algorithm during the preliminary analysis of information for the portfolio.

KEYWORDS

Doctor's portfolio, continuous professional development of doctors, intelligent algorithms of information analysis, co-referencing of medical information, secondary nomination of medical terms, cognitive structures

1. Introduction

Today, the portfolio mechanism is used in the world not only as a method of qualitative transfer of knowledge, but most importantly, as a mechanism for qualitative evaluation of the acquired knowledge! Therefore, the problems of improving the accounting mechanism, assessing the quality of acquired knowledge in the system of continuous medical education are discussed very widely. Among the large number of such evaluation mechanisms, the application of electronic Portfolio technology is more relevant today both for the formalization of knowledge and competences, their monitoring, and for the holistic assessment of the quality of education in the system of continuous medical education.[1–4]

Currently, portfolio is a web technology that allows a doctor to record his educational activities and professional successes throughout the entire period of professional development. However, the logic of the approach to assessing the need to acquire knowledge for different medical professions based on average indicators, the same approach to measuring very different competencies in different medical specialties, level the peculiarities of professional activity.

In addition, the decentralization of the system of acquiring knowledge, the emergence of new centers for the training of doctors caused the problem of comparing and evaluating educational programs, and the problem of understanding texts of high complexity became obvious even for qualified representatives of the medical profession. The secondary nomination of medical terms[5] is one of the obvious difficulties in the integration of educational activities. Accordingly, the priority task is to ensure the co-reference of clinical information related to the comparison of data obtained from different sources for their inclusion in the portfolio.[6,7]

The purpose of this work is to justify the prerequisites for creating a new type of portfolio by using the methods of intellectual analysis of multidimensional information, as well as the formation of models and cognitive structures, which should become a central event in person-oriented active learning.

Today, problems of improving the mechanism of accounting and assessing the quality of acquired knowledge in the system of continuous medical education are constantly being discussed. Among the large number of such assessment mechanisms, the application of the "electronic portfolio" technology is more relevant both for the formalization of knowledge and competences, their monitoring, and for the integral assessment of the quality of education in the system of continuous medical education.

We consider the portfolio as a personal professional-oriented technology for the authentic assessment of the educational results and industrial successes of a doctor, an effective means of quantifying educational and professional growth in his continuous professional development.

The portfolio can be used to demonstrate the competences acquired by the doctor in various areas of medical education, including interdisciplinary and transdisciplinary ones, assimilation by the doctor of certain non-clinical skills, and apply them in his work. In many countries around the world, the use of the portfolio is mandatory in order to obtain and confirm the qualification of a doctor.[3,4]

The amount of characteristics that probably characterize the training and work of a doctor is counted in the hundreds. In addition, the quantitative assessment of educational procedures is often complicated by the unstructured nature of the information. Therefore, it is quite difficult to give a valid integral description of the activity of a specialist. But in today's world, working with big data is the norm. For this, there are many powerful methods of information processing, among which the methods of Intelligent Data Analysis (IDA) occupy an important place.

It is generally accepted that IDA is a process in which, with the help of mathematical and computational algorithms, raw data is structured and various regularities in these data are formulated or recognized.[8] For the portfolio, the integral assessment of professional growth is quite difficult, first of all, as a result of the lack of formalization of ways to improve qualifications.

An excessively large number of options for obtaining knowledge, a non-unified system of assessment of knowledge and competences are also important. It is logical to justify typical models of professional development (which is constantly replenished). Then the task of integral characteristics of the professional development of the individual is reduced to a simpler problem of pattern recognition. But even in this version, we have quite a large number of difficulties associated with qualitative (not quantitative!) indicators, erroneous information, etc. Given that currently none of the algorithms can provide a valid assessment of the professional growth of a specialist, we suggest using *ensembles of algorithms*, that is, a combination of several algorithms that learn simultaneously and correct each other's mistakes. This approach is often used in order to strengthen the "positive qualities" of individual algorithms, which can work poorly on their own, but in a group - give a good result.[9]

We use two classic approaches of ensembles – bagging (in which the basic algorithms are representatives of the same family, they are trained in parallel and almost independently of each other, and the final results are only aggregated) and boosting (the models are no longer trained separately from each other, but each of the next one manages the errors of the previous one, meaning that if one weak algorithm failed to detect any pattern in the data because it was too hard for it, then the next model should do it.

Additional coefficients were used to improve recognition results. One of them was called the *prioritization coefficient*, and the second is the *coefficient for determining the content equivalence* of various thematic professional development courses. The importance of determining priorities is due to the need to select from the total amount of information about the doctor that needs priority attention for use in the assessment. In this way, it will be possible to assess the specialist's competence as soon as possible, as well as the need to correct the trajectory of his professional development. Of course, higher priority information should be processed first.

Considering that the portfolio is a classic information technology as well as the fact that the educational process requires continuous evaluation and management of the quality of service provision, we used the SLA (Service Level Agreement) strategy. Today, this technology is used and has a high level of trust by millions of people around the world.[10] The SLA system of post-graduate training of doctors and pharmacists consists of four parts: a glossary, a brief description of the systems, the roles of process participants and the results of the technological learning process. The characteristics of the SLA action are important - territorial, temporal and functional. The basis in this direction is the ratio of the characteristics of the subject area and the volume of the educational material, the depth of study of the educational materials, the level of evidence of the information and the reliability of the connections between the educational elements.

Taking into account that a change in priority can affect the target performance metrics, the calculated priorities must be well understood by all participants in the process. Changes in priorities should occur in parallel with the dynamics of solving the problem and with the appearance of new information that can radically change the initial formulation of the task.

The *coefficient of determining the content equivalence* of various thematic courses for the improvement of specialists is extremely important in the conditions of difficulties in solving problems of uncertainty in the choice of training courses, certification of training course providers, training quality control systems, etc. There is a need for informal comprehensive measurement and the simultaneous use of various assessment tools, methods of multidimensional analysis and special methods of integration of scores on various quantitative and qualitative scales. It is necessary to determine which content obtained as a qualification improvement is equivalent, and which is essential for obtaining a qualitative change in the doctor's qualification.

The content equivalence of various thematic courses is also specified using the VEN (Vital Essential Non-essential) strategy, which is a prescriptive segmentation of the "necessity" of educational elements. This strategy is quite often used in medicine and production; especially together with ABC analysis.[11]

The result of ABC analysis is the grouping of educational objects according to the level of influence of the overall result on the level of knowledge transfer. In our problem, this strategy will allow us to determine priority sections of educational content in accordance with international standards and answer the question: "Which educational courses should be considered for inclusion (exclusion) in the list of educational services?". The use of ABC analysis will allow ranking the range of training courses according to various parameters. It is possible to organize educational content providers, the quality of knowledge transfer, and knowledge management systems and other educational resources in this way.

The last module in decision-making regarding the professional growth of an individual is the analysis of texts of high complexity. The essence of the problem lies in the huge multiple nomination of pathology terms (the emergence of terminological "traps", or co-referentiality), which are further complicated by the terminological discrepancy of pathological conditions in diploma and postgraduate education.

In the absence of special rules, coreference names will be used differently in different sections of text systems. Accordingly, there is a problem of unification and standardization of input, search, extraction, storage and exchange of data both in relation to the completion of educational procedures and the assessment of professional competences in various clinical situations.

The integration of medical information systems into a single space involves the provision of common referentiality of clinical information related to the comparison of data obtained from different sources. But the great complexity of determining the semantic compatibility of various text structures during computer processing of electronic records does not allow today to effectively solve the problem of the correct record in the portfolio of a specialist.

To solve the coreference problem, it is recommended to use special software systems.[12] The most common approach is to build a model of the mentioned pairs.[13] The classifier first identifies all the mentioned pairs that are co-referential. These pairs are then grouped into chains using clustering methods such as the nearest neighbor method. The second option is the mention rating model.[14] With this method, the task is reformulated as a ranking function rather than a classification function. All candidate mention antecedents are ranked to determine which candidate is the most likely.

Another well-known method is the use of the previous tree model.[15] In this case, a graph is constructed from the document, where nodes are mentions, and arcs are connections between pairs of mentions that are coreference candidates. Then coreference chains are modeled as latent trees on the graph.

To solve the coreference problem, we used ontological models (graph models) and correspondence rules. Cognitive load is reduced if taxonomic techniques are used, which organize the terms in the controlled dictionary into a hierarchy. The main purpose of taxonomy is to create an ontological structure for human understanding of information and its integration from different sources.

Thus, decision-making during the preliminary analysis of information for inclusion in the portfolio can be presented in the form of an algorithm (Fig. 1).

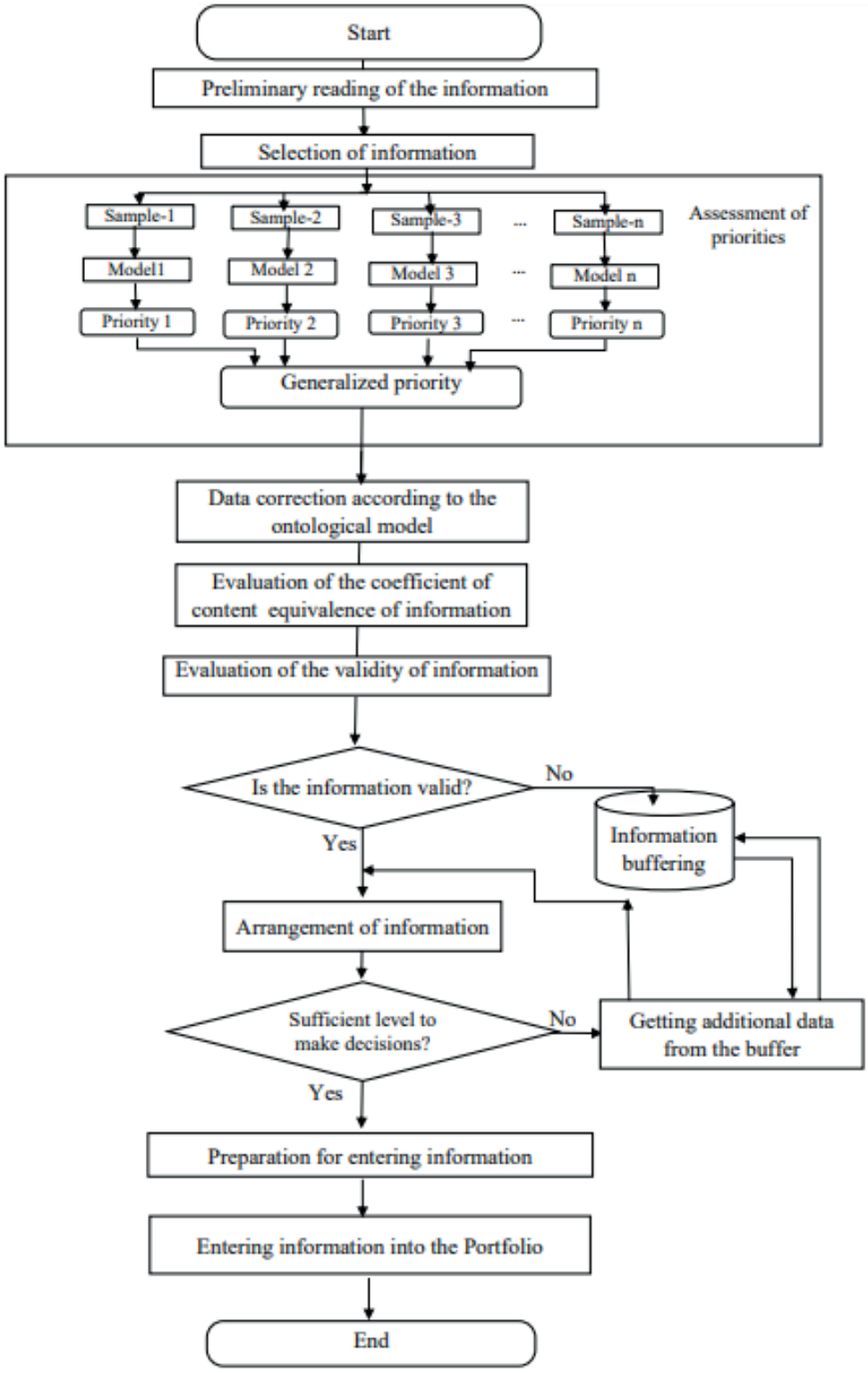

Figure 1: Algorithm for decision-making during the preliminary analysis of information for inclusion in the portfolio.

As can be seen from the algorithm, the main processes in the processing of information before entering it into the doctor's portfolio are the evaluation of its validity and the organization of data. The biggest difficulties, the solution of which forms the basis of the validity of information, are the issues of secondary nomination and coreference of data. Note that we are talking only about the preliminary analysis of information before entering it into the portfolio. Even greater problems await the researcher at the second stage of information processing. They are related to the content analysis of the accumulated data and will be discussed in future publications. At the same time, what has been said emphasizes the complexity of creating a portfolio, and taking into account the importance of introducing the portfolio mechanism into health care practice – the need for a comprehensive discussion of the technology of its creation.

2. Conclusions

1. The portfolio of doctors and pharmacists is extremely important for the functioning of the system of continuous professional development of the relevant specialists. It helps organize lifelong learning and allows you to keep all evidence of learning and professional activity.

2. Taking into account the importance of the portfolio, it should be capable of standardized and formally calculated evaluation.

3. The entry of information into the portfolio is preceded by the determination of the semantic equivalence of the available information regarding the acquisition and assimilation of knowledge.

4. It is proposed to use ensembles of approaches to the analysis and arrangement of information in the portfolio - a combination of several algorithms that function simultaneously and provide an opportunity to correct possible errors.

5. The proposed decision-making algorithm for the analysis of preliminary information for the portfolio.

References and notes

[1] De Swardt M, Jenkins LS, Von Pressentin KB, Mash R. Implementing and evaluating an e-portfolio for postgraduate family medicine training in the Western Cape, South Africa. *BMC Med Educ*. 2019;19(1). doi:10.1186/s12909-019-1692-x

[2] Sukhanova O. Logics and basic principles for evaluation of quality of doctors' continuing professional development using the electronic portfolio. *Regional Innovations Medical Science*. 2021;2:11-14.

[3] Tochel C, Haig A, Hesketh A, et al. The effectiveness of portfolios for post-graduate Assessment and Education: BEME Guide No 12. *Medical Teacher*. 2009;31(4):299-318. doi:10.1080/01421590902883056.

[4] Van Tartwijk J, Driessen EW. Portfolios for assessment and learning: AMEE Guide No. 45. *Medical Teacher*. 2009;31(9):790-801. doi:10.1080/01421590903139201.

[5] Tsisar N. System-forming potential of secondary medical names. *Bulletin of the Lviv Polytechnic National University (Series "Problems of Ukrainian Terminology)*.2010;675:54–56

[6] Mintser OP, Babintseva LYu, Sukhanova OO. The logic of concordance of indicators of knowledge assimilation in the doctor's portfolio. *Medical Informatics and Engineering.* 2022;1-2(57-58):56-63. doi:10.11603/mie.1996-1960.2022.1-2.13112

[7] Mintser OP, Popova MA, Ladychuk OK, Koshova SP. Taxonomy usage for postgraduate content improvement. *Medical Informatics and Engineering.* 2020:3(51):33-40. doi: 10.11603/mie.1996-1960.2020.3.11604

[8] Kotova EE. Intellectual data analysis in the educational process. *2017 XX IEEE International Conference on Soft Computing and Measurements* (SCM). 2017. doi:10.1109/SCM.2017.7970714

[9] Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol.* 2019;19(1). DOI: 10.1186/s12874-019-0681-4.

[10] England R. *Mastering ITIL. (Translation from English)*. Livebook. 2011.

[11] Schawel C, Billing F. Top 100 Management Tools. 2011. doi:10.1007/978-3-8349-6605-6

[12] Ferracane E, Marshall L, Wallace BC, Erk K. Leveraging coreference to identify arms in medical abstracts: An experimental study. Seventh International Workshop on Health Text Mining and Information Analysis (LOUHI). 2016. 86-95. Austin

[13] Soon WM, Lim DC, Ng HT. A Machine Learning Approach to Coreference Resolution of Noun Phrases. *Computational Linguistics.* 2001;4(27):521-544.

[14] Denis P, Baldridge J. Specialized Models and Ranking for Coreference Resolution. In Proceedings of the 2008 Conference on Empirical. *Methods in Natural Language Processing.* Honolulu. Hawaii. 2008. 660-669. doi:10.3115/1613715.1613797

[15] Thorsten J, Chun-Nam JY. Learning structural SVMs with latent variables. In Proceedings of the 26th Annual International Conference on Machine Learning, ACM. 2009. New York