



---

## Design and Filling of the Vocabulary of the Lexical Selection of Potential Entrants' Motivational Intentions

Anna Shilinh<sup>1</sup>, Pavlo Zheznych<sup>2</sup>

<sup>1</sup>Department of Social Communication and Information Activities, Lviv Polytechnic National University, 12 S. Bandery str., Lviv, 79000, Ukraine, AnnaShiling@gmail.com

<sup>2</sup>Department of Social Communication and Information Activities, Lviv Polytechnic National University, 12 S. Bandery str., Lviv, 79000, Ukraine, pavlo.i.zheznych@lpnu.ua

---

**Abstract** The aim of this article is to describe the vocabulary's structure of the identified motivational intentions of potential entrants. Formation or refinement of the motivational intentions of potential customers of educational services takes place during the entry campaign. The motivational intentions of potential entrants differ depending to the period of the entry campaign. Lexical selection of motivational intentions of potential entrants, discovered in communicative activities, makes it possible to clearly formulate / adjust the activity of the higher education institution during the entry campaign. For a rational allocation of time for the reaction of the higher education institution to the participants' posts of the educational web-communities for each period of the entry campaign, it is necessary to structure the identified motivational intentions. As a result of the work, the vocabulary's structure of the motivational intentions of potential entrants is described, taking into account the peculiarities of the formation of informational content of educational web-communities. Each such motivational intent, which is included in the vocabulary, is characterized by a weight coefficient, which shows the relevance of a some post in a particular period of the entry campaign. The statistical indicator of TF-IDF was used to evaluate the importance of informational content of the vocabulary of motivational intentions. The results of the research are applied and can be used to effectively plan the activity of the higher education institution in relation to the periods of the entry campaign. According to the results of the research, recommendations were given to the use of the vocabulary of motivational intentions for effective planning of educational services provided by higher education institutions.

**Keywords** higher education institution, vocabulary, motivational intent, weighting factor, importance of content

---

### 1. Introduction

The decision-making process for potential entrants in choosing a future specialty during an entry campaign depends on various factors: economic, social, domestic, physiological and psychological. Upon request, the entrance preparation process (specialty selection, examinations, submission of an e-application for entry, etc.) prior to entry to a higher education institution (HEI) of the chosen specialty takes place in a virtual environment (electronic cabinets and official web resources of HEI). That is why, the search for the future entrant's information is also happening in the web-space. Official information about the specialties of preparation for the HEI is available on the HEI's website in the catalog of educational programs, and the search for more specific, more subjective, information calls is the need to create thematic discussions in the educational web-communities. By formulating requests for clarification or explanation of information, potential entrants invest in their information needs and motivational intentions, that generate these needs. The selection of motivational intentions from the text-posting of thematic discussions of potential entrants enables HEI to more clearly plan or adjust its activities during the entry campaign and to determine the directions and methods of direct influence on



the decision-making process by consumers of educational services. An important aspect of the formation of the motivational intentions vocabulary is the determination of the importance of its content, since it enables accurate and rapid response to the information needs of participants in communicative activities.

## **2. Related Works**

Today, virtual space is the primary source of communicative activity, rather than the medium for passive informatization [1, 2]. Research and analysis of the communicative activities of participants in web communities is the subject of work [3]. In particular, research [4] analyzes communicative activities in specialized web communities, taking into account the peculiarities of planning for the provision of educational services. The tendency of decision-making by entrants during the selection of the HEI and future specialties is analyzed in works [5, 6]. A large amount of informational content of educational web-communities puts the demand for its intellectual processing and analysis [7]. Methods of intelligent Data Mining and Text Mining are considered in [8]. In particular, works [9, 10] contain an analysis of methods for solving the problem of extracting information from texts. The linguistic foundations of extracting terminological units from scientific and technical texts are the subject of the study [11]. The structure of the specialized vocabulary of the lexical representation of the topic extraction from scientific literature for competency management in [12]. The methods for determining the importance of words in the web space and the criteria for the quality of web pages are the subject of research [13, 14]. In [15], the choice of the statistical indicator TF-IDF for the evaluation of the importance of words in the notes and determining the relevance of this post is grounded. But none of the studies are considering the possibility of integrating the information obtained to effectively plan educational services by higher education institutions.

## **3. The aim and objectives of the study**

The analysis of thematic discussions text posts of educational web-communities requires extracting motivational intent from entrants, that is, words or phrases that identify them, and determine the relevance and relevance of the specific period of the entry campaign. That is why, the aim of this article is to create a motivational intentions vocabulary of potential entrants identified in the discussions of educational web-communities and to evaluate the importance of the motivational intentions found to determine the relevance of the post and its timely response to it. The objectives of the work are to describe the vocabulary structure of the motivational intentions of potential entrants and to determine the criteria for evaluation of the importance of its information content.

## **4. Description of the vocabulary structure of the motivational intentions of potential entrants**

The motivational intentions of potential entrants are an indicator of their interest in choosing a future specialty and higher education institution [5]. They change during the entry campaign and depend on its period. The formation of the vocabulary of motivational intentions allows you to structure the findings in relation to the periods of the entry campaign and determine the areas of possible influence on future entrants, analyzing their needs, which determine these motivational intentions.

The basis of the formation of vocabulary lexical selection is the comparison of parts of the text messages of potential entrants with the elements of the set of keywords for each period of the entry campaign. The basis for selecting the motivational intentions of potential entrants is to identify syntactical units that indicate the needs of these users in relation to each period of the entry campaign. This role in the posts of participants in communicative activities are key words. This set of keywords for each period determines the main directions of activity of the higher education institution. Description of the structure of the vocabulary of motivational intentions of potential entrants is presented in Figure 1.



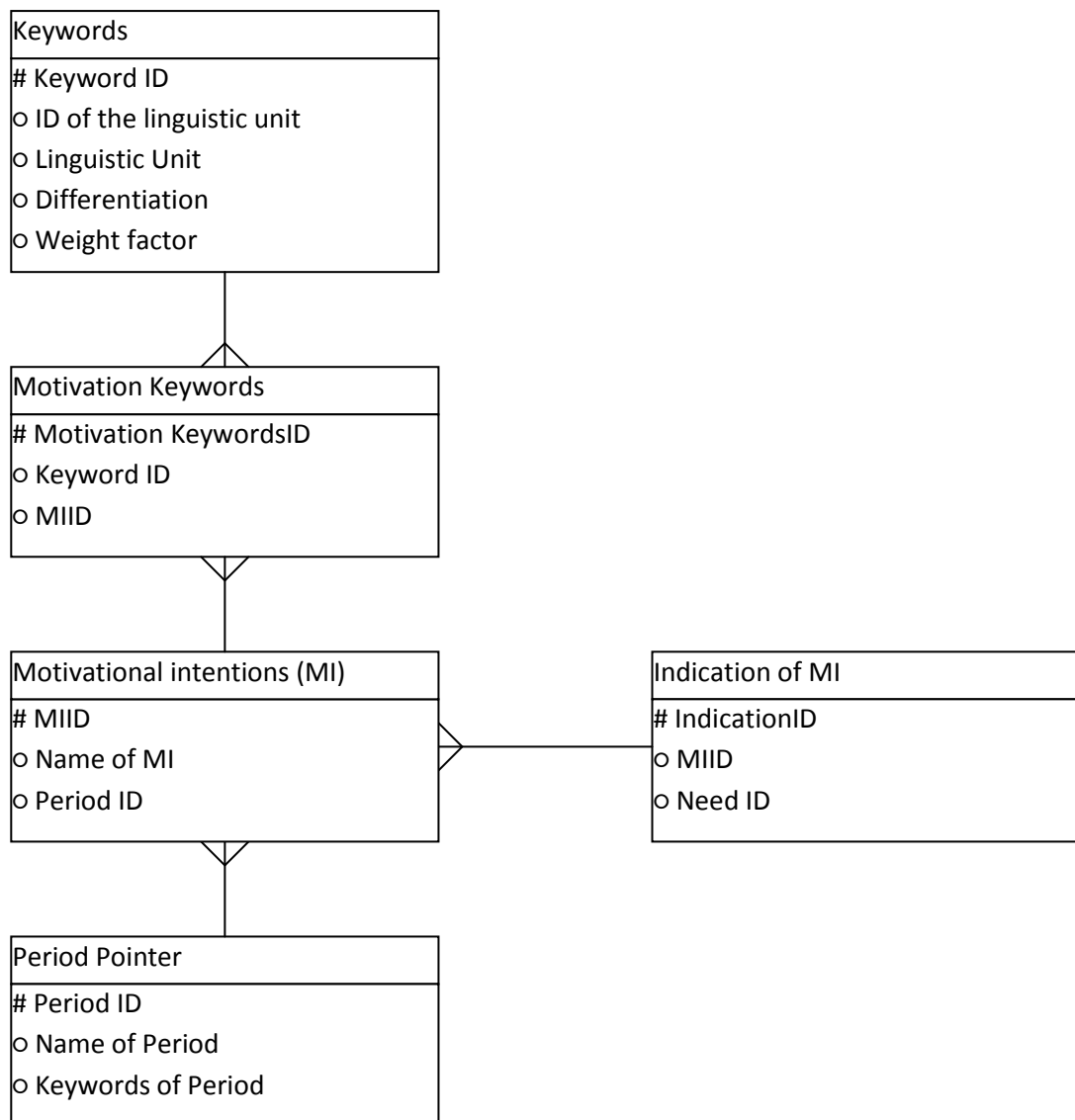


Figure 1: Structure of the vocabulary database of motivational intentions of potential entrants

The «Motivational intentions (MI)» relation contains data motivational intentions of potential entrant that match your keywords («Keyword ID» attribute) specified period of утрекн campaign («Period ID» attribute).

The «Indication of MI» relation includes signs of motivational intentions of entrants. Since motivational intentions are provoked by the needs of consumers of educational services, the components of motivational intentions include the «Need ID» attribute is the user needs. The «Need ID» attribute contains information about the name of the need. For example, if the motivational intentions of the discussion participants relate to the "entry rules", then the need for the user is "joining the HEI", as the designation of the direct process, and the "certificates", "exam", etc., as the designation of the information needs.

The "Keywords" relation contains information on the possible lexical-semantic forms of one or more keyword labels that characterize the needs of users of the educational web-communities relative to a certain period of the entry campaign. For example, the information need of participants in communicative activities is the name of this need (obtaining information on specialties, or entry, the need to share experiences, etc.). The "Differentiation" attribute shows the possibility of applying a specific value to the "Linguistic Unit" attribute to replace the motivational intentions of potential entrants. Depending on the possibility of differentiation, it takes values on the interval [0, 1]. For example, for motivational intentions that indicate the information need for lexical-semantic designations of this need, are the names of educational qualification programs or the names of



the HEI structural units. The "Weight factor" attribute shows the importance factor of the motivational intentions of the subject discussions detected in the posts.

The «Motivation Keywords» relation ensures the relationship of the set of keywords to the set of motivational intentions of potential entrants.

The "Period Pointer" relation contains data language constructs that indirectly indicate motivational intentions for a certain period of the entry campaign. That is, those who do not call the period, but point to it. The connection between motivational intentions due to the specific need of a potential entrant is carried out through keywords that relate to a certain period of the entry campaign.

### 5. Results of forming the motivational intentions vocabulary of potential entrants from the posts of the educational web-communities

The use of the proposed algorithm for the formation of the motivational intentions vocabulary of potential entrants (for example, the information content of a specialized educational web-forum for entrants) for the 2013-2018 entry campaigns is shown in Table 1.

**Table 1:** Motivational intentions relative to a certain period of entry campaigns 2013-2018

The period of the entry campaign	Examples of motivational intentions	Examples of keywords	Number of identified keywords
I period	«How to entry», «how to study», «which institution carries out training»	«entry», «education», «institution of higher education»	176
II period	«which certificates», «which examinations are to be compiled», «is there a creative competition»	«Certificate», «examinations», «competition»	51
III period	«when to submit an electronic application», «what privileges are foreseen», «how additional points are taken into account»	«electronic application», «privileges», «additional point»	119
IV period	«when to submit originals», «where it is possible to review the rating list enrollment», «when it is possible to draw up a contract»	«original», «rating», «list», «зарахування», «contract»	462
V period	«whether there is a swimming pool», «when the beginning of classes», «when the settlement in the hostel», «how the scholarship is awarded»	«swimming pool», «classes», «hostel»«scholarship»	42

Consequently, the motivational intentions of potential entrants are the expression of their actual needs in relation to a certain period of the entry campaign. Timely impact on the main areas of the formation of the needs of potential entrants in relation to each period of the entry campaign, which are determined by the motivational intentions vocabulary, plays an important role in the decision-making process of the future student.

### 6. Evaluation of the keywords importance that describing the motivational intent of potential applicants

The motivational intentions of potential entrants are determined by their needs and characterized by the keywords of a certain motivational intention within the specific period of the entry campaign [16]. For a timely and qualitative response from the HEI, its representatives must determine the relevance of the post not only in the timeframes of the period of the entry campaign, but also in evaluation the keywords importance in it. All keywords, that characterize motivational intentions in the posts of potential entrants, determine the relevance of the post regarding other thematic discussions.

The TF-IDF statistic was used to evaluation the importance of identified keywords. The TF-IDF indicator is used in text analysis and information search tasks.



The importance of keywords proportional to the number of occurrences of the word in the post, and is inversely proportional to the frequency of the use of the word in other posts of the thematic discussion:

$$TF - IDF = TF \cdot IDF \quad (1)$$

where  $TF = \frac{n_i}{\sum_k n_k}$  - the ratio of the number of occurrences of a particular keyword to the total number of words in the post,  $IDF = \log_2 \frac{|D|}{|d_i \ni t_i|}$  - the inversion of the frequency with which the keyword occurs in the postings of the thematic discussion.

TF measures the importance of a keyword within the selected post. Using IDF reduces the weight of widely used words. The TF-IDF will receive high-frequency keywords in the post and low frequency in other posts in the thematic discussion.

The results of calculating the importance of keywords, that characterize the entrants' motivational intentions in the vocabulary, according to a certain period of the entrance campaign in 2018, are presented in Table 2.

**Table 2:** Calculation of the TF-IDF

The period of the entry campaign	Motivational intent of the period	TF	IDF	TF-IDF
I period	«entry»	0.058	0.766	0.045
II period	«examination»	0.15	1	0.15
III period	«application»	0.1	1	0.1
IV period	«original»	0.125	1	0.125
V period	«deanery»	0.14	1	0.14

In accordance with the procedure for evaluation of the keywords importance, it is worth noting that the calculation does not notice the mention of a particular specialty, if in the post, she is mentioned by name. That is why the list of key words should include a list of all the specialties that are under preparation for a higher education institution, the names of all structural units of this HEI (institutes, faculties, departments), for a successful and timely response to a post in the thematic discussion. So, the TF-IDF indicator was used to evaluate the importance of vocabulary information filling. This indicator allows you to identify relevant posts that contain the motivational intentions of potential entrants in relation to a certain period of the entry campaign. Since the speed and quality of the reaction of the HEI directly depends on the timeliness of identifying the actual post, then evaluating the keywords importance of the motivational intentions is a necessary indicator for solving this problem.

## 7. Conclusions

Consequently, the proposed motivational intentions vocabulary of potential entrants aims to structure the main aspects for the decision making process, selected from the information content of thematic discussions of educational web-communities. Since the presence of motivational intentions testifies to the need for a participant in communicative activities, this poses the task of the HEI in a timely and qualitative way to react to them. The clear actions of the higher education institution are the basis for effective planning of educational services. The formation of such a vocabulary allows you to identify relevant presentations and to influence the basic needs of educational services users in relation to each period of the entry campaign. Each motivational intent is characterized by a weighting factor. To determine the motivational intentions importance, the TF-IDF was used, which evaluates keywords within a post and reduces the importance of widespread words. But the calculation of the information content importance of the dictionary does not take into account the keywords if they contain the specific name of the structural unit of the higher education institution. That is why the list of keywords needs to be expanded taking into account the names of the specialties, institutes and departments of this HEI.

## Reference:

- [1]. Vus, V., Albota, S., Dobrovolska, V. (2019). The Analysis of Online Communities as Platforms for Informational Influences. *Journal of Scientific and Engineering Research*, 6(2), 72-78.



- [2]. Burnett, G. Information Exchange in Virtual Communities: A Comparative Study [Electronic resource]. *Journal of Computer-Mediated Communication*. Vol. 9, Issue 2. – Access mode: (<http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2004.tb00286.x/full>.)
- [3]. Zhang, J., Yu, J.X., Hou, J. (2006). *Web Communities: Analysis and Construction*, Springer, Heidelberg, 187 p.
- [4]. Shiling, A, Zhezhnych, P.(2012).Analysis of the user behavior specialized forums university entrants. *Proceedings of the VIIIth International Conference on Computer Science and Information Technologies (CSIT'2012)*. Lviv: Publisher Lviv Polytechnic, 202.
- [5]. Connelly, G. Halliday, J. (2001). Reasons for choosing a further education: the views of 700 new entrants. *Journal of Vocational Education & Training*, Vol. 53(2), 181-192.
- [6]. Mikhno, N. Sorokina, L. (2017). Trend Analysis of the Educational Choice of University Entrants as Tool to Improve the Quality of Educational Service. *International Journal of Information and Communication Technologies in Education*, Vol. 6(3), 36-39.
- [7]. Yi, L., Liu, B. (2003). Eliminating Noisy Information in Web Pages for Data Mining. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2003)*, Washington DC, USA, 296 — 305.
- [8]. Santos, A., Guillen, P.G. Villa, E., Serradilla, F. (2012). Semantic Construction of Univocal Language. *Information Theories and Applications*, Vol. 19, Number 3, 211—215.
- [9]. Corcho, O., Lopez, M., Gomez-Perez, A. (2003). Methodologies, tools, and languages for building ontologies. Where is their meeting point? *Data & Knowledge Engineering*. Vol. 46, 41-64.
- [10]. Buitelaar, P., Cimiano, P., Magnini, B. (2005). Ontology Learning from Texts: An Overview. In *Ontology Learning from Text: Methods, Evaluation and Applications*, Vol. 123, 234-265.
- [11]. Simperl, E., Mocho, M. (2010). Achieving Maturity: the State of Practice in Ontology Engineering. In *International Journal of Computer Science and Applications, Technomathematics Research Foundation*, Vol. 7, № 1, 45-65.
- [12]. Buileaar, P., Eigner, T. (2008). Topic extraction from scientific literature for competency management. In *The 7th International Semantic Web Conference PICKME*, Germany, 55-67.
- [13]. Page, L., Brin, S., Terry, W. The PageRank citation ranking: Bringing order to the Web [Electronic resource] – Access mode: (<http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>).
- [14]. Song, R., Liu, H., Wen, J.-R., Ma, W.-Y. (2004). Learning Block Importance Models for Web Pages. *Proceedings of the 13th international conference on World Wide Web*, New York, USA, 203—211.
- [15]. Jones, K. S. (2004). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* : Vol. 60, № 5, 493-502.
- [16]. Zhezhnych, P., Shilinh, A. (2019) Linguistic Analysis of User Motivations of Information Content for University Entrant's Web-forum./ *International Journal of Computing*, 18(1). 67-74.

