Information Technology for Constructing Individual Educational Trajectories Based on Latent-Semantic Analysis of Motivational Letters and Professional Achievements of Students

Tetiana Kovaliuk[1]'[0000-0002-1383-1589]', Nataliya Kobets[2]'[0000-0003-4266-9741]' and Victoria Dvornyk[3]'[0000-0002-2019-1786]'

[1]Taras Shevchenko National University of Kyiv, 64/13, Volodymyrska Str, Kyiv, 01601 Ukraine,
[2]Borys Grinchenko Kyiv University, 18/2 Bulvarno-KudriavskaStr, Kyiv, Ukraine, 04053
[3]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Prospect Peremohy, Kyiv, Ukraine, 03056
tetyana.kovalyuk@gmail.com, nmkobets@gmail.com, vikadvornik@gmail.com

Abstract. The article considers the possibility of constructing of individual educational trajectories by students, the significance of self-determining the way of self-development. The relevance of the work is determined by introducing the paradigm of student learning. The construction of individual educational trajectory is reduced to the problems of analyzing students' motivation and determining academic disciplines which will constitute the variable part of student's individual curriculum. Students' motivations were analyzed by the applying content analysis to students ' motivational letters. As a result, the data on students' priority fields of knowledge and activity directions were obtained. Determination of academic disciplines is carried out on the basis of Latent semantic analysis (LSA). The results of the content analysis of the students' motivational letters, the results of the students' career focus survey and the results of tests on students' current knowledge levels are the input of the LSA. The output of Latent semantic analysis is a list of knowledge areas relevant for the student's profession. Ontologies of subject areas are built based on the list of knowledge areas. Ontology of subject areas is proposed to be built in the form of a thesaurus, describing the categorically-conceptual apparatus of the subject area. Determination of the key words of the subject area topics chosen by the student is carried out by using the Latent Dirichlet allocation. Defined keywords are the criteria for selecting the disciplines which will constitute the student's individual curriculum. Approbation of proposed information technology has shown its reliability.

Keywords: individual educational trajectory, motivation letter, content analysis, latent semantic analysis, Latent Dirichlet allocation, student.

1 Introduction

Target indicators in the development of the education sector by 2030 are determined among the global goals of sustainable development: to ensure inclusive and equitable quality education and to promote lifelong learning opportunities for all [1]. In terms of human development, education significantly expands human capabilities, as it positively influences different aspects of life, health, public and political activity, access to knowledge and ability to use it both at work and in everyday life, raising children, etc. The education level is one of the three components of the human development index together with GDP per capita and average life expectancy. According to [2] the opening of each new university raises GDP per capita of the country by an average of 0.05% and increases the volume of the region's economy, where this university is by 0.4%.

The successful development strategy in the context of knowledge economy depends on the high quality of human capital. In Ukraine, 282 universities produce 1.322 mil. of Bachelors and Masters. However, the quality of their training does not meet the requirements of the industry. According to Global Competitiveness Report 2019 [3] the rating of Ukraine by the "Skills of current workforce" indicator is 53, by "Quality of vocational training" - 65, the indicator "Ease of finding skilled employees" is 53. Thus, the problem of higher education quality improvement is urgent.

One of the key topics discussed at the World Economic Forum-2020 in Davos was Reskilling Revolution [4]. This initiative aims to prepare people for the professions of the future in perspective up to 2030. According to experts 133 mil. new jobs will emerge in economies of the majority of countries by 2022, and by 2030, it is necessary to retrain 1 billion people, because 42% of the basic skills will change in the nearest future. The biggest challenge is to motivate people to retrain to acquire new knowledge and skills relevant to the knowledge economy.

The need of transition to the knowledge economy challenges the education system to improve the quality of educational processes for mass training of specialists. This means that the training of specialists, which in terms of their number and quality, social, personal and professional competencies would meet the requirements of the industry, is a relevant problem for universities in Ukraine.

One way to improve the quality of higher education is to implement the paradigm of student-centered learning [5]. The basis of the student-centered learning is the idea of providing the students with maximized chances to get their first job in the labor market, an increase of their "value" for employers, i.e. their suitability for employment. The nature of the paradigm of student-centered learning is to personalize learning. The model of individual approach in learning aims to sustain the effectiveness of the learning process and is carried out through the implementation of individual educational trajectories. Individual educational trajectory is a personal way of realizing the personality potential of a student and is formed taking into account their active, cognitive, creative, communicative and other abilities.

Students will be able to advance along individual trajectory in all educational spheres if they are given the ability to determine the individual content of academic disciplines, to set their own goals studying a particular topic, to choose optimal forms and pace of

learning, to apply methods of training that meet their personal features. Thus, implementation of student-centered learning and constructing individual educational trajectories taking into account motivation, psychological, professional and personal qualities of a student is a modern trend of education development.

According to the "On amendments to certain legislative acts of Ukraine on improvement of educational activity in the sphere of Higher Education" law [6], upon admission to study to obtain a bachelor's degree the entrants, besides the results of external Independent Evaluation of knowledge and creative competitions will have to submit a motivation letter to the universities. The prospective students should describe their abilities, principles, skills, attitudes, education, professional experience, etc. Their motivation letters should demonstrate positive thinking, active life position, and far-reaching plans for the future. Motivational letters of students can be taken into consideration for constructing an individual educational trajectory.

2 Analysis of Related Research and Publications

Many scientific works of domestic and foreign researchers are devoted to the problems of student-centered learning, individual educational trajectories and the choice of forms, methods and techniques of organizing the educational process in higher education institutions [7, 8]. The concept of student-centered education is the basis of the Law of Ukraine "On Higher Education" [9], in particular on issues of academic mobility, free choice of curriculum by students, etc. According to Tuning Guide [10], student-centered learning aims to increase the independence and critical ability of the future professional through results-oriented activities. Student-centered learning strategies implementation, with the aim of developing the 21st century skills of self-directed and lifelong learning in students are discussed in [11].

A wide range of different views and methodological approaches to personal educational learning trajectories are presented in [12, 13, 14, 15, 16, 17]. A common opinion among all authors is that an individually oriented educational trajectory always implies a student's freedom to choose various elements of the educational process and organizational forms of educational activity.

A study of student motivation for choosing to continue studies and selecting majors, and the impact on motivation of lecturer and student behavior was held, and the results are presented in [18]. Authors [19] believe that the motivated student has the inner strength to learn, to discover and capitalize on capabilities, to improve academic performance and to adapt to the demands of the university context.

The structural and semantic model for realization of individual educational trajectories within the technological platform is elaborated in [20]. Modern research in the field of education is focused on the methodological foundations of the construction and operation of information systems in the management of the educational institution and application of multimedia technologies in education. However, there are practically no works addressing the problems of automated analysis of students' motivational letters, the requirements of employers and the construction of individual educational trajectories based on them.

3   Problem Statement

The purpose of this work is to substantiate the feasibility of automated processing of students' motivation letters to construct individual educational trajectories. To achieve this goal, it is necessary to solve the following problems:

- performing latent semantic analysis of motivation letters;
- identifying areas of expertise that are targeted by students' motivations;
- identifying the professional competencies of a student;
- developing automated procedures for composing individual curriculums.

The information technology under development will provide identification of students' personal qualities based on psychological tests, personality and character tests, determination of the student's career focus based on job aptitude tests to clarify students' psychological, professional and personal qualities.

4   Technological Process of Student Individual Curriculum Formation

In this paper the possibility of practical application of the educational process individualization is addressed on the example of IT education. Therefore, further it will be focused on IT industry and IT education topics. The limitation of the system is the formation of an individual educational trajectory for one semester or one academic year. The technology of constructing individual educational trajectory of student is shown on Figure 1.

Students, stakeholders (representatives of companies interested in individual training of students according to the requirements of employers) and tutors (curators or guarantors of educational program) are participants of the individual educational trajectory creation process.

Employers' requirements may be present in the program system as a result of stakeholder surveys, data collection from open information sources (job sites, European e-Competence Framework, Skills Framework for the Information Age, ACM Competency Model For Graduate (Undergraduate) Degree Programs etc.), higher education standards and educational programs of higher education institutions.

To take into account the students' expectations, motivations, psychological, professional and personal qualities it is necessary to determine the base level of knowledge and their potential abilities. For this purpose, a student can undergo a number of tests, including a psychological test, a leadership test, the existence and level of professional competence development test. As a result of testing, students may be recommended professions that best suit their abilities. However, a student may specify a profession at will, disregarding the program recommendations.

After selecting a profession, a webpage of motivation defining is available to students. Students are encouraged to upload, or input into corresponding field his motivation letter or essay.
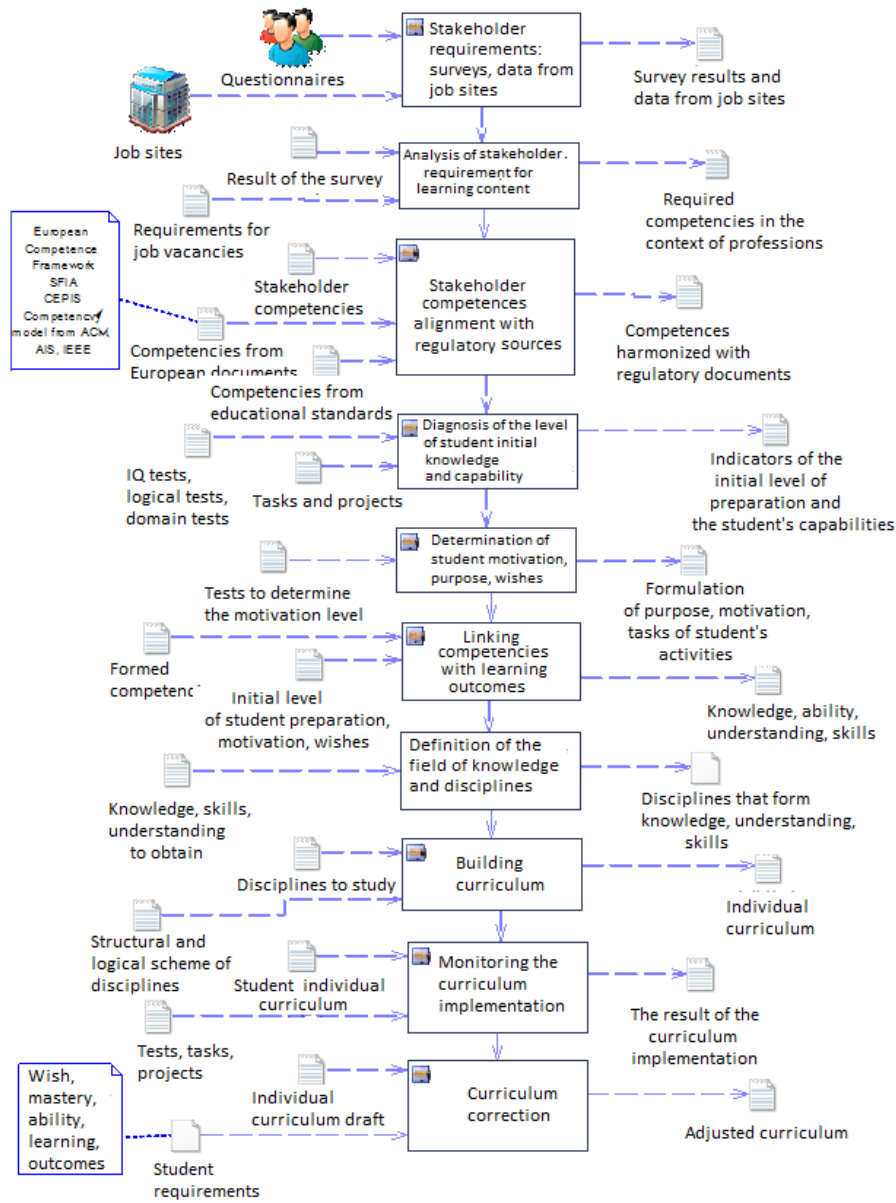
Fig. 1. Technology of constructing individual educational trajectory of students

The information system will perform lexical-semantic analysis of the text or document and will determine the level of motivation at which the student's activities will be

successful. The criteria for evaluating motivational letters include analysis of the following indicators:

- clarity and consistency of the wording of the letter;
- rationality and concreteness of expressing the ideas;
- ability to summarize, analyze and evaluate own professional experience and outline the prospects for further professional development;
- understanding of modern trends in education development, existing problems and ways to solve them;
- ability to see the development horizons of the educational sector, in particular IT education and IT industry;
- IT and language literacy, knowledge of professional terminology, individual author's style;
- acquired competencies and student's professional guidelines.

The results of the analysis of motivational letters, in particular, the competencies possessed by the student, soft skills and hard skills that the student wants to obtain, their vision of the individual educational route, the results of previous tests are the basis for the formation of recommended subjects list. This data is given to a tutor, who can approve the formed list or adjust it. For this purpose, the tutor looks at the key words of the fields of knowledge for the profession that applicant has chosen, defines the disciplines from the knowledge areas offered by the system, number of credits and the form of control. According to the requirements of the provisions on organization of educational process, the number of disciplines and variety of control forms in higher education institution are determined. The formed list of disciplines is written to the database and is submitted for students to acquaint themselves.

The student should approve the list of disciplines developed based on their motivational letter and test results. Individual educational trajectory contains the branch of knowledge for which it was built, student data (name and surname), personal trajectory name, applicant's course number (year of study) for which the trajectory was formed and a list of disciplines that student will study, number of credits and forms of control.

5 Motivation Letter as a Formalized Text Model

One of the simplest methods for presenting texts in natural language is the bag-of-words approach [21, 22]. The bag-of-words text model is a multi-set of words that comprise the text. A word is the main object of the bag-of-words model. A word is characterized by the only attribute - the frequency of its occurrence in the text. The bag-of-words text model takes into account only the number of words in the source text, ignoring the word order in the document and morphological forms of word representation.

The generalization of the formal bag-of-words text model is the bag-of-terms model. A term is a character expression of the object of the formal model of a language, a system, a text. By formal definition, the term is a symbolic expression $t(X_1, X_2,..., X_n)$

where t – term name, and $X_1, X_2,..., X_n$ – structured or the simplest terms. Terms may contain free variables (parameters) values of which are clearly determined by a certain object in accordance with the semantic rules of the language. Obviously, the term value is determined by the set values of its free variables. In logic-mathematical calculus, the term is an analogue of the subject or supplement in natural language. In the bag-of-terms model, any symbolic expressions of the text, including punctuation, can be considered as terms. For each word in the bag-of-words set, a certain "weight" may be specified. Thus, the text model is a set of "word-weight" pairs. To determine the weight of words, the following methods are used:

- binary method (BI). Only the presence or absence of certain terms in the document is determined.
- number of words in the document. More weight is gained by documents that have more words;
- frequency of word inclusion in a document (TF), it is calculated as the ratio of the number of a word inclusions to the total number of words in the text.
- logarithm of word inclusion frequency (LOGTF). The weight of the document is determined by the expression $1 + \log(TF)$, where TF determines the frequency of a in the text;
- Inverse Document Frequency (IDF). The option is an inversion of the word inclusion in the documents frequency.

The bag-of-words text model is the basis for the Vector Space Model, where a text is represented as vectors from one common for the whole text collection vector space. Text or document in the vector model is considered an unordered set of terms. Terms are the words that make up the text, and other text elements that are important in the model (numbers, punctuation, special designations, acronyms, etc.).

The main entities in constructing natural language vector representations are a token t, a document d, a corpus C and a vocabulary V. An elementary unit of text, usually a word, is called a token. A document is an unordered set of tokens. The document can be a motivational letter, an article, a proposal, a review. A case is all available documents combined. The vocabulary is a set of all unique tokens found in the case. Each token is matched with a unique index from 1 to $|V|$. Thus, after building the dictionary, two representations are available: index $\rightarrow$ token, token $\rightarrow$ Index.

A vector, which is a model representation of text in the vector space, is formed by arranging the weights of all terms (including those that are not in a particular text). The dimension of this vector, as well as the dimension of space, is equal to the number of different terms in the entire collection, and is the same for all texts (documents) of the collection.

Formally:

$$d_j = (w_{1j}, w_{2j},..., w_{nj}), j \in J,\qquad(1)$$

where d_j – vector view of the j-th document (motivation letter), w_{ij} – weight of the i-th term in the j-th document, $n = |V|$ – total number of different terms in all documents of the collection, i.e. the volume of the vocabulary.

To define the vector model fully, it is necessary to specify exactly how the term weight in the document will be found. The downside of the bag-of-words model is ignoring the semantic connections between words. To overcome the limitations of the bag-of-words model, the text is represented using latent topics. Such representations are formed by the "documents-terms" matrix factorization [23, 24]. The most popular method among this class of approaches is Latent semantic analysis (LSA), which uses singular decomposition as a factorization.

6 Latent Semantic Analysis Method and Latent Dirichlet Allocation

The problem of constructing an individual student curriculum is decomposed into the tasks of determining students' motivations and determining the list of disciplines that will be included in the individual curriculum of a student. The task of determining the list of disciplines is decomposed on the sub-tasks of determining the list of areas of knowledge for the student's chosen profession and identifying keywords on the subjects of selected areas of knowledge. Let us consider the task of checking students' motivation when choosing a specialty and constructing an individual educational trajectory. As a source information, we use motivation letters, essays and tests on job aptitude in the IT industry with open-ended questions. To filter, rubricate, cluster documents, search for answers to questions, automatically annotate documents, search for similar documents and duplicates and for automated assessment of the quality of answers to open-ended questions it is proposed to use the Latent semantic analysis (LSA) [25, 26] and the TF-IDF scheme [27,28].

6.1 Latent Semantic Analysis Method to Test Students' Motivation

Suppose there is a collection of n documents that are student's motivation letters and m different terms derived from the IT term dictionary. The term refers to a separate word or phrase. To apply the latent-semantic analysis model, a $m \times n$ "terms-documents" matrix X has to be built, the x_{ij} values of which contain the term weights coefficients $t_i, i = \overline{1, m}$ in the document $d_j, j = \overline{1, n}$. The rows of the matrix correspond to terms "weighted" by a metric, such as TF-IDF or based on entropy. The matrix columns correspond to the documents; the elements are the frequencies of use of terms in documents. The resulting "terms-documents" matrix is a spatial-vector model of textual information in natural language and contains input data for the Latent semantic analysis method. Whether the input text in natural language relates to the IT industry and the topic of the text has to be determined.

Frequency matrixes can be sparse and noisy, so it is advisable to perform reduction and singular value decomposition of the matrix. Reduction is necessary to extract semantic connections that relate to the user's topic of interest (subject area). The reduction of the matrix is discarding part of its columns and rows. The "extra" rows of the matrix, corresponding to the terms of other topics, are cut off. Then the "extra" columns of the matrix are cut off, corresponding to the documents of the negative set. As a result, we get a reduced "terms-documents" matrix, columns of which correspond to the texts of the topic in question, and the rows correspond to its keywords.

The second transformation of the "terms-documents" matrix consists of its singular value decomposition, which results in the $m \times n, m > n$ matrix X being presented as a multiplication of three matrixes $X = USV^T$, where U and V^T are respectively $m \times r$ and $n \times r$ orthogonal matrixes, S is a $r \times r$ diagonal matrix containing its own values. Each of its r own values corresponds to one of the components in the collection of documents, and indicates how relevant this component is throughout the collection.

Own values are sorted diagonally in matrix S descending, as a result the first own values are associated with the most important components. This allows to reduce the least important components to the amount $k \leq r$ by removing the relevant rows and columns in the matrixes. Reducing the matrix removes "noise" in data that can be for example terms or groups that appear only in a few documents and are poorly associated with others. The singular value decomposition of the matrix replaces it with a matrix of the same dimension, but of a smaller rank, in which only the most important information is retained.

The "terms-documents" matrix, reduced and cleared of noise, is used to build a "terms-terms" semantic relationship matrix. Since in the "terms-documents" matrix each term is a vector-row, the semantic relationship between any two terms can be interpreted as the distance between the respective terms vectors, using any known measures of proximity or distance [28]. In this work we use a cosine measure:

$$\cos(\overline{T_i}, \overline{T_j}) = \frac{\overline{T_i} \times \overline{T_j}}{|\overline{T_i}| \times |\overline{T_j}|}, \tag{2}$$

where $\overline{T_i}, \overline{T_j}$ - vector-strings of a "terms-documents" matrix corresponding to i-th and j-th terms respectively. Values close to 1 mean similar terms (documents), values close to 0, represent their heterogeneity.

6.2 Latent Dirichlet Allocation Application for Determining Disciplines and Topic Modeling

To determine the affiliation of a given document (text) to a particular topic Latent Dirichlet allocation is used, according to which each document is considered as a set of different topics.

Suppose a collection of text documents D is given. Every document $d \in D$ is a sequence of words $W_d = (w_1, ..., w_{n_d})$ from the vocabulary V, where n_d – the length

of the document d. The number of topics K is set. It is assumed that each document may relate to one or more topics. Topics differ from each other by different frequency of words use. Every topic $t \in T$ is described by an unknown distribution $p(w|t)$ on a set of words $w \in W$. Every document $d \in D$ is described by an unknown distribution $p(t|d)$ on a set of topics $t \in T$. The probability model of the "term-document" pair (w,d) can be given as (3):

$$p(w,d) = \sum_{t \in T} p(d)p(w|t)p(t|d) \qquad (3)$$

With additional assumptions:

- document vectors $\theta_d = (p(t|d):t \in T)$ are generated by the same probability distribution on normalized $|T|$-dimensional vectors, which belongs to the parametric family of Dirichlet allocations $Dir(\theta, \alpha), \alpha \in R^{|T|}$;
- vectors of topics $\beta_k = (p(w|t):w \in V)$ are generated by the same probabilistic distribution on rationed dimensional vectors $|V|$ belonging to the parametric family of Dirichlet allocations $Dir(\theta, \beta), \beta \in R^{|V|}$.

6.3 Describing Algorithms

Let us consider the Latent-semantic analysis algorithm.

Step 0. Exclude stop words from analyzed documents. Stop words are the words that do not carry semantic meaning; they are all the conjunctives, particles, prepositions and a multitude of other words. Go to the step 1.

Step 1. Filter numbers, individual letters and punctuation marks from analyzed documents. Go to step two.

Step 2: Perform a stemming operation to get the base word above all words from the documents. At the stemming stage, the endings and suffixes of words are discarded to determine the unchanging part of the word. The use of lexicographical dictionaries and morphological analysis at the lemmatization stage will lead to defining the main form (lemma) of each word in a sentence. Go to step three.

Step 3. Make a frequency matrix of indexed words. In this matrix, the columns correspond to documents and rows are indexed words. In each cell of the matrix, there must be a specified number of repetitions of words in the corresponding document. Go to step four.

Step 4. The resulting matrix should be normalized with the TF-IDF matrix normalization method. Go to step five.

Step 5. Carry out two-dimensional singular decomposition of the obtained matrix. Go to step six.

Step 6. Discard the last columns of the matrix U and the last rows of the matrix V^T leaving only the first two columns. This corresponds to the x, y coordinates of each

word for the matrix U and the x, y coordinates for each document in the matrix V^T.

Step 7. Prepare output data as nested lists of x, y coordinates for two columns of the words matrix U and two rows of the documents matrix V^T. Go to step eight.

Step 8. Compare the coordinates of the words of a given dictionary (in motivation checking case) or terms of the areas of knowledge (in the defining disciplines case) with known documents using a cosine distance.

Let us consider the Latent Dirichlet allocation algorithm.

Step 0. Assess the average number of topics in document K and the average number of keywords in subject V. Go to step one.

Step 1. Simulate the distribution of topics on documents and words by Dirichlet allocation with the parameters: K-dimensional vector α (the smaller α, the fewer topics in the document), V-dimensional vector β (the less β, the fewer words that characterize the topic). Usually all coordinates of vectors α and β are given the same. Go to step two.

Step 2: For each document, generate topics probability by Dirichlet allocation with parameter α: go to step three.

Step 3. For each topic generate probability of Dirichlet distribution words with parameter α. Go to step four.

Step 4. For each item in the document choose a random topic according to the generated probabilities, choose a random word, according to the words probabilities in the topic.

7 Software Implementation of Information Technology

The IDE Visual Studio 2017 with the Python Tools for Visual Studio plugin, as well as the SQLite database, were used to create the software product. The application architecture is based on the Model View Controller (MVC) pattern. Django framework was used. The spaCy module as Python packages and the English grammar were used to implement the Natural Language Processing algorithm. The textacy library was used to implement the algorithm for extracting semi-structured expressions. Web application testing was performed on the corporate server nemo.asu.kpi.ua.

The reviewed algorithms were implemented in the module of formation of fields of knowledge towards which students are motivated according to their motivational letters of the Web application. The diagram of the developed software components is presented in Figure 2.

Web application enables the student to take a psychological test, a test of leadership and test on the professional competencies formation. Student can choose the IT profession recommended by the program system, to test the motivations to mastering knowledge and skills in selected profession, to construct the individual trajectory of training, which consists of selected disciplines of individual curriculum.

Information technology uses an ontology of the educational program, which is built with the help of domain experts and presented in Figure 3. This ontology

simultaneously serves as a "filter" for selecting fragments of motivation letters that potentially contain information that is processed by the described algorithm.
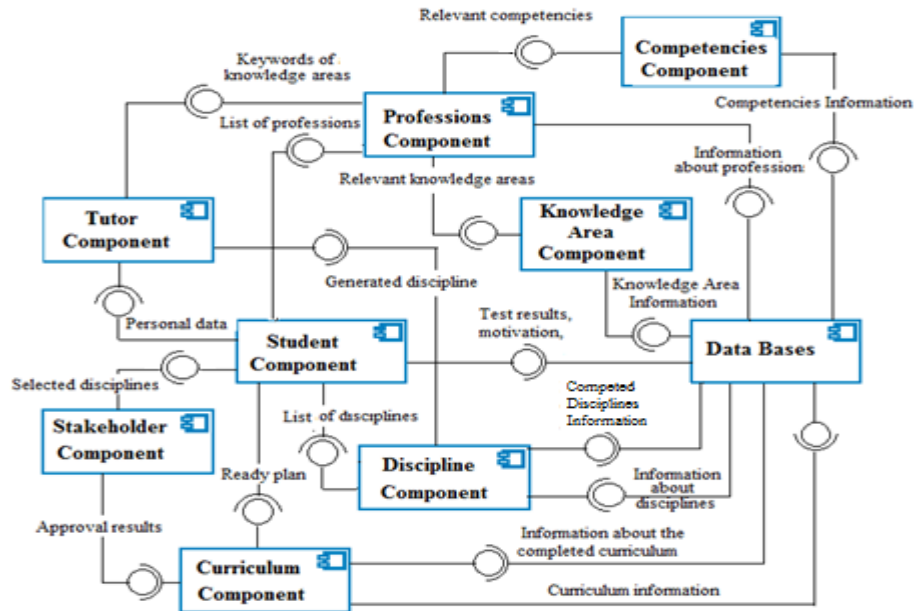

Fig. 2. Component diagram of information technology of constructing individual educational trajectory
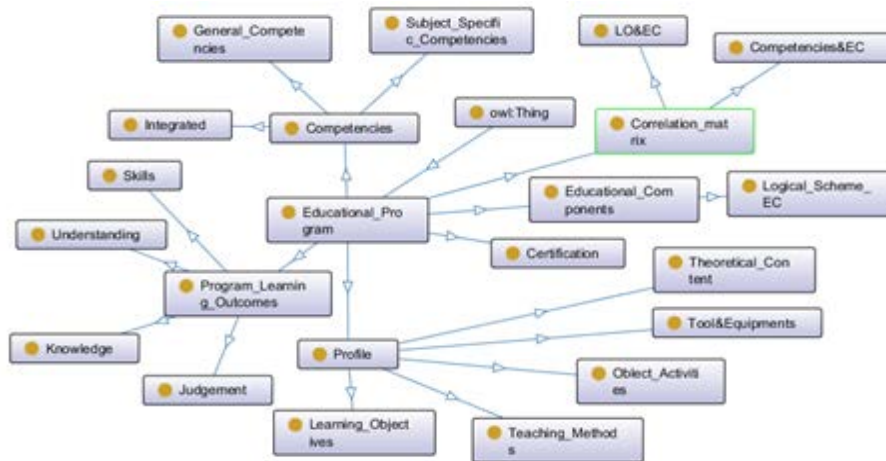

Fig. 3. Ontology Educational Program

8   Conclusion

The paper examines the problem of constructing an individual educational trajectory, which is represented by two tasks: determining the student's level of motivation and determining the list of disciplines based on the analysis of motivation letter and of the psychological, professional and personal qualities of the student tests results.

The Latent semantic analysis method and TF-IDF scheme for the implementation of the classifier were chosen to solve the problem of testing students' motivation and determining the list of disciplines. The Latent Dirichlet allocation method was used to determine the list of areas of expertise for the student's chosen profession and to identify keywords in selected areas of expertise.

Software implementation of information technology has shown the effectiveness of the methods considered.

Further development of research is directed to the ontology and thesaurus automatic building, and accounting the Ukrainian grammar.

References

1. Sustainable development goals knowledge platform: https://sustainabledevelopment.un.org/sdg4
2. Valeroa, A., Van Reenen, J.: The economic impact of universities: Evidence from across the globe. Economics of Education Review, vol. 68, 53–67 (2019)
3. Schwab, K.: Global Competitiveness Report 2019. World Economic Forum: http://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf
4. Towards a Reskilling Revolution: A Future of Jobs for All. World Economic Forum: http://www3.weforum.org/docs/WEF_FOW_Reskilling_Revolution.pdf
5. Bezanilla, M., Wagenaar, R., González Ferreras, J.: Tuning Educational Structures in Europe. Final Report. Pilot project-Phase 1. Learning outcomes: Competences. University de Deusto (2003)
6. Law of Ukraine "On amendments to certain legislative acts of Ukraine on improvement of educational activity in the sphere of Higher Education" of December 18, 2019: https://zakon.rada.gov.ua/laws/show/392-20?find=1&text#Text (2009)
7. Kovaliuk, T., Pasichnyk, V., Kunanets, N., Veretennik, N.: Professional competency management of IT professionals to industry requirements based on cognitive cards. Information Technologies and Learning Tools, vol. 64, no 2, 253–264 (2018)
8. Kovaliuk, T., Pasichnyk, V., Kunanets, N.: Modeling the development of higher education based on the competence approach and personality-oriented educational trajectories. Information technologies and Learning Tools, vol. 61, no 5, 245–260 (2017)
9. Law of Ukraine "On Higher Education" no 1556-VII of 18.03.2020: https://zakon.rada.gov.ua/laws/show/1556-18#Text
10. Lokhoff, J.: A Tuning Guide to Formulating Degree Programme Profiles. Bilbao, Groningen and The Hague (2010)
11. Jacobs, G. M., Toh-Heng, H. L.: Small steps towards student-centred learning. In P. Mandal (Ed.), Proceedings of the international conference on managing the Asian century pp. 55–64. Singapore: Springer (2013). doi: 10.1007/978-981-4560-61-0_7
12. Zakharchenko, V. M., Lugovy, V. I., Rashkevich, Yu. M., Talanova, Zh. V.: Development of educational programs. Guidelines. State Enterprise: Priorities, Kyiv (2014)

13. Mukhametzyanova, F. G., Zabirova, R. V.: Designing an individual educational trajectory and route for a university student – future bachelor. Kazan pedagogical magazine. https://cyberleninka.ru/article/n/proektirovanie-individualnoy-obrazovatelnoy-traektorii-i-marshruta-studenta-vuza-buduschego-bakalavra (2015)

14. Maslow, A. H.: Motivation and Personality. Harper & Row, Publishers, Inc. (1970)

15. Harkusha, Yu., Sheludko, S.: Individual educational trajectories in professional training process. In: Vzdelávanie a spoločnosť IV. Medzinárodný nekonferenčný zborník. Prešovská univerzita v Prešove, pp.199-204 (2019)

16. Pozharkova, I. N., Noskova, E. E., Troyak, E. Yu.: Formation individual educational trajectory as a component of practice-oriented learning environment. Pedagogical image. 3 (40), 179–192 (2018)

17. Ibatullin, R. R., Anisimova, E. S.: Construction of individual educational trajectory of students based on e-learning. In: IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), Baku, pp. 1–4 (2016)

18. Doghonadze N.: Undergraduate students' motivation. In: The fifth International Conference on Education and New Learning Technologies, Barcelona, Spain, 1-3 July, pp. 4116–4128 (2018)

19. Ferreira, M., Cardoso, A., Abrantes, J. L.: Motivation and Relationship of the Student with the School as Factors Involved in the Perceived Learning. In: International Conference on Education and Educational Psychology (ICEEPSY 2011), 29, pp. 1707–1714 (2011)

20. Zeer, E. F., Streltsov. A. V.: Technological Platform for Realization of Students' Individual Educational Trajectories in a Vocational School. International Electronic Journal of Mathematics Education, vol. 11, no 7, pp. 2639-2650 (2016)

21. Nugumanova, A. B., Immortal, I. A., Petsina, P., Baiburin. E. M.: Enrichment of the bag-of-words model with semantic connections for improving the quality of classification of texts of the subject. Software products and systems, 2 (114), 89–99 (2016)

22. Wallach H.M.: Topic modeling: beyond bag-of-words. In: Proceedings of the Twenty-Third International Conference on Machine learning (ICML 2006), Pittsburgh, Pennsylvania, USA, pp. 977-984 (2006)

23. Weinan, E, Yajun Zhou.: A Mathematical Model for Linguistic Universals. https://arxiv.org/pdf/1907.12293v1.pdf

24. Mashechkin, I., Petrovsky, M., Popov, D., Tsarev, D.: Automatic text summarization using latent semantic analysis. Programming and Computer Software, vol. 37, no 6, 299-305 (2011)

25. Thomo. A.: Latent Semantic Analysis. https://www.engr.uvic.ca/~seng474/svd.pdf (2009).

26. Chen S., Ma B., Zhang K.: On the similarity metric and the distance metric. Theoretical Computer Science, vol. 410, no. 24, pp. 2365-2376 (2009)

27. Blei, D. M., Ng A. Y., Jordan M. I.: Latent Dirichlet Allocation. Journal of Machine Learning Research, vol. 3, 993-1022 (2003)

28. Palmer D. D.: Text preprocessing. Handbook of Natural Language Processing, Second Editionm, Chapman and Hall/CRC. 9–30 (2010)