The 8th International Scientific Conference
eLearning and software for Education
Bucharest, April 26-27, 2012
10.5682/2066-026X-12-125

ALGORITHM FOR DECOMPOSITION OF LEARNING CONTENT

Mitko MITEV, Anatoliy ANTONOV, Ivaylo PENEV
Department of Automation and Computing, Technical university of Varna, Studentska Str. 1, Varna, Bulgaria
E-mail: mmmitev@abv.bg, antonov@eurorisksystems.com, ivailopenev@yahoo.com

***Abstract**: The paper presents development and application of an algorithm for decomposition of learning content into separate units for the purposes of distant learning. In accordance with the purpose the limiting conditions and the aims of the decomposition are defined. The computing complexity of the algorithm is formed. The algorithm's programming realization could be used both in the development and estimation of learning plans and programs or to be included as an extension of existing distant learning platforms. The algorithm is partially applied into the development of new specialty "Software and internet technologies".*

***Keywords:** notions, links, directed graph, subgraphs, decomposition, learning units, distant learning*

I. INTRODUCTION

One of the main problems [1] into the development and preparation of courses for distant learning is dividing the learning content into separate, logically connected learning units. The volume of the units is of great importance. It depends on lots of factors. The most important are degree of learning, depending on the individual qualities of the person and on the rate of learning, presence of regulated period for carrying out the learning process and not on last place current status of the person[2]. Lots of these factors are hard to be predicted and in most cases cannot be estimated. Therefore, in methodical point of view, the system of learning is required to be adapted according to the status and the possibilities of the learning person.

Independent from the dynamics of learning, the formation of separate units, specific to each person, has to satisfy definite requirements:

- The units have to include possible maximum set of strongly connected notions or facts.
- The units should be relatively independent, i.e. to be in a complete state. The logical links between the units should be minimal.
- The links between the separate learning units must follow the time direction.

These requirements are common and do not present the specific features of the specific learning process and the individual characteristics of the person.

II. THEORETICAL FORMULATION

The development of the theoretical formulation is based on the following prerequisites [3]:

1. Each learning unit is consisted of set of mutually connected notions, supported by facts and illustrative material.

2. Notion, on which another notion is based, is called *basic*, and the connected one – *sequence*.
3. The link between two notions is called *causal link*. It is directed from the basic notion to its sequence.
4. The link's direction could follow the direction of time. In this case the basic notion is situated before its sequence. Such a link has *straight direction.*
5. Otherwise, if the sequence precedes the basic notion, the link has *reverse direction*.
6. Each solution, which orders notions in a way that the links in straight direction dominate in a definite meaning, is called *feasible solution*.
7. From methodical point of view "*distance*" in time between the basic notion and its sequence is necessary to be minimal. By indirection the distance is in proportion with the estimated notions, situated after the basic notion and before its sequence on the time axis.
8. *Rational* solution is the one from the set of feasible solution, in which the sum of the distances between the basic notions and their sequences has local (in particular case - global) extreme.

This theoretical formulation considers with priority the structure of learning content. All the following researches and obtained results are based only on its structure. The project solutions of forming learning units consider only the identified notions and the causal links existing between them.

Consequently the division of the learning content into units must not be assumed as finite, but only as *first approximation*.

III. FORMALIZATION

Let we, according to [4] denote with *X* the set of notions, included into the learning material. The power of the set is *N* and is defined by the number of notions, i.e. $|X| = N$.

The set of the established causal links is denoted by *U*, where $|U| \leq N^2$.

According to [5] the following unambiguous relations exist:

$$(\forall x_i \in X)(\exists w_i \in W)\&(\forall w_i \in W)(\exists x_i \in X) \rightarrow (X \leftrightarrow W) \quad (1)$$

$$(\forall u_{ij} \in U)(\exists f_{ij} \in F)\&(\forall f_{ij} \in F)(\exists u_{ij} \in U) \rightarrow (U \leftrightarrow F) \quad (2)$$

, where w_i is weighted coefficient of $x_i \in X$ and relatively f_{ij} in accordance with $u_{ij} \in U$, $i, j \in \{1,2, \ldots.. N\}$. The weighted coefficients are discussed in [6].

The incident links between the elements of *X* and *U* are formed by three-seated predicate:

$P(x_i, u_{ij}, x_j) = true$ – if $x_i \in X$ is basic notion, $x_j \in X$ is its sequence and relatively *false*, if there is no causal link between the notions.

According to this formalization the structure of the learning content could be presented as finite, oriented graph *G(X,U,W,F,P)* with weighted vertices and arcs.

The sum of the weights of all vertices is denoted by W_B

$$W_B = \sum_{\forall w_i \in W} w_i \quad (3)$$

and in the general case presents expert estimation for the complexity of the analyzed learning content.

Let us assume that $W_0 \ll W_B$ presents recommended limit value for the volume of each learning unit. Consequently the expected number of learning units, denoted by *M* could be calculated.

$$M = \left] \frac{W_B}{W_0} \right[\quad (4)$$

with precision to the first greater or equal integer number.

IV. DECOMPOSITION ALGORITHM

Let us assume that an a priori solution $r_0 \in R$ is given from the set of all possible solutions, presenting the graph *G* as set of sub graphs,

$$G = \{g_1, g_2, \ldots . g_m, \ldots g_M\} \quad (5)$$

where

$$(g_m \subset G) \rightarrow g_m(X_m U_{mm} W_m F_{mm} P) \tag{6}$$

or

$$X = \{X_1, X_2 \ldots\ldots X_m \ldots\ldots X_M\} \tag{7}$$

From the unambiguous relation $X \leftrightarrow W$ follows

$$(\forall X_m \subset X)\left(\exists \sum_{\forall x_i \in X_m} w_i \le W_0\right) \rightarrow (\Delta w_m = W_0 - \sum_{\forall x_i \in X_m} w_i) \tag{8}$$

The difference Δw_m, which occurs by objective reasons presents estimation for "*consolidation*" of each learning unit to the limit volume.

Referring to Δw_m the following considerations could be done:

1. $\Delta w_m \rightarrow min$ is recommended, which could be assumed as learning unit with minimal rest of time.
2. It is desirable that

$$(\forall k, m \in \{1, M\}) \rightarrow (\Delta w_k \cong \Delta w_m) \tag{9}$$

i.e. the learning units to be equally consolidated in time.

3. A particular case exists, in which following the previous two requirements is not recommended. For example, when Δw_m as time is used about questions, answers or instructions for self preparation.

Due to the discrete values of $w_i \in W$ the value *M*, determined by (4) is possible to be increased.

The following relations are assumed in the decomposition algorithm

$$M \rightarrow min \tag{10}$$

$$\sum_{\forall X_m \in X} \Delta w_m \rightarrow min \tag{11}$$

To some degree the two criteria are dependent as minimizing the sum of objectively obtained differences is prerequisite for decreasing the number of learning units. This dependence is not valid in the general case.

According to (5) and in analogy with (7) the decomposition solution $r_0 \in R$ determines the following sub sets of graph arcs:

$$(k \neq m) \rightarrow U = \{U_{m,m}, U_{k,m}, U_{m,k}\} \tag{12}$$

, where

$$(k < m) \rightarrow (U_{k,m} \equiv \overrightarrow{U_{k,m}})\&(U_{m,k} \equiv \overleftarrow{U_{m,k}}) \tag{13}$$

$U_{m,m}$ - set of *inner* arcs

$\overrightarrow{U_{k,m}}$ - set of *outer* arcs in straight direction

$\overleftarrow{U_{m,k}}$ - set of *outer* arcs in reverse direction

Taking into account the unambiguous relation (2), the following sums of weights of the inner and outer arcs could be calculated:

$$(\forall g_m \in G) \rightarrow (s_{mm} = \sum_{\forall u_{i,j} \in U_{mm}} f_{i,j}) \tag{14}$$

$$(\forall g_k g_m \in G)(k < m) \rightarrow \left(\overrightarrow{s_{km}} = \sum_{\forall x_i \in X_k \& \forall x_j \in X_m} f_{ij}\right) \tag{15}$$

Analogically the sum of weights for the outer arcs in reverse direction (case *k* > *m*) is calculated.

In summary

$$S_{mm} = \sum_{\forall m \in \{1 \ldots M\}} s_{mm} \rightarrow max \tag{16}$$

$$\overrightarrow{S_{km}} = \sum_{\forall k < m} \overrightarrow{s_{km}} \rightarrow min \tag{17}$$

$$\overleftarrow{S_{mk}} = \sum_{\forall k < m} \overleftarrow{s_{mk}} \rightarrow min \tag{18}$$

or

$$\forall (k \neq m) \rightarrow \left(\overrightarrow{S_{km}} + \overleftarrow{S_{mk}}\right) \rightarrow min \tag{19}$$

From the set of solutions *R* this one is recommended, which in conditions (17) and (18) determines $\overrightarrow{S_{km}} > \overleftarrow{S_{mk}}$, i.e. as the whole sum from the weights of the outer arcs to be minimal and in equal conditions the sum of the weights in straight direction to dominate over the sum from the weights in reverse direction.

Assuming this treatment the algorithm for the decomposition of the graph *G* has to satisfy the following purposes:

1. Minimizing the number of learning units and the consolidation connected to this according to the determined limitation.
2. Each learning unit has to include the tightly connected notions.
3. The connections between the separate units must be decreased to the possible minimum.
4. The solution has to determine dominating links in straight direction in comparison to the links in reverse direction.

In preliminary estimation the development of decomposition algorithm, satisfying to maximum extent the defined purposes is hard to be realized in practice due to the following factors:

- Some of the purposes are mutually contradictory, for example the consolidation of the learning units can cause increase of the outer links.
- Each change of the places of the vertices with the purpose of decreasing the arcs in straight direction can cause inaccessible increase of the arcs in reverse direction.
- The tight connection of the graph can cause insignificant improvement of the intermediate results and so on.

Independent from the complexity of the problem, caused by the presence of the limiting condition W_0 and the connected requirements (10) and (11), as well as the defined purposes (16), (17) and (18) is proposed algorithm for decomposition from the class of rational solutions. The algorithm is based on operation $\alpha(x_i, x_j)$, which mutually changes the places of the two vertices. Each preorder of the vertices determines new solution.

In summary the algorithm includes execution of the following points:

1. Introducing the graph *G(X,U,W,F,P)* with quality estimation of the elements of the sets and relatively the limiting condition W_0.
2. Locating the vertices of the graph in linear sequence in accordance with the assigned number (index).
3. Executing the algorithm [3] for reordering the vertices in dependence to the choice of one of the three criteria:

- Local minimum of the sum of the arcs in straight and reverse direction.
- Local minimum of the sum of the arcs in straight direction.
- Local minimum of the sum of the arcs in reverse direction.

4. Finding initial solution $r_0 \in R$ (5) in case of satisfied restriction W_0, defined *M* and relatively:
$$\{U_{mm}\}; \{U_{km}\}; \{U_{mk}\}; \{\Delta w_m\} \text{ for } \forall m, k \in \{1,2,\dots M\}$$
5. For each *m* ∈ {*1,2,…M-1*} are explored the adjacent pairs the sub graphs $g_m g_{m+1} \in G$ in sequence. For this purpose:
   - 5.1. The sub sets $X_m \subset X$ и $X_{m+1} \subset X$ are sorted according to the weights of the included vertices, as follows: X_m in descending and X_{m+1} in ascending order.
   - 5.2. According to Δw_m and w_i of the vertex $x_i \in X_m$ is determined the range of the vertices of X_{m+1}, which do not violate the limiting condition W_0
   - 5.3. Analogically, according to Δw_{m+1} and w_{i+1} of the vertex $x_j \in X_{m+1}$ is determined the range of the vertices of X_m, which do not violate limiting condition.
   - 5.4. For each pair of vertices $x_i \in X_m$ and $x_j \in X_{m+1}$ from the obtained ranges are calculated the sums of the weights of arcs in straight, reverse and in both directions (according to preliminary chosen rule). The calculations are performed before and after applying the operation $\alpha(x_i, x_j)$.
   - 5.5. If applying the operation $\alpha(x_i, x_j)$ satisfies the defined purposes the result is accepted as a new solution. Otherwise the next pair of vertices is processed.
6. The algorithm stops when the state *m* = *M* is reached.

V. ESTIMATION OF THE CALCULATING COMPLEXITY

For heuristic determination of the calculating complexity of the proposed decomposition algorithm the following is assumed:

- Equal distribution of the power of the sub sets of the vertices, forming the subgraphs:

$$(\forall m \in \{1, M\})\left(n_m \cong \frac{N}{M}\right) \tag{20}$$

- If the structure interpretation of the learning material approaches to full graph, then the number of the inner arcs is Mn_m^2 and the outer arcs relatively $N^2 - Mn_m^2$,
- or $(\forall x_i \in X_m)$ the number of inner arcs is $2n_m$ with precision to 1 and the outer arcs relatively $2(N - n_m)$
- Sorting each sub graph by the method *Bubble sort* has extent of n_m^2 for comparisons and changes.
- The total number of analyzed pairs adjacent sub graphs is *M-1* in case of one iteration or totally *M(M-1)*.

Consequently the total quantity of the operations comparison will be in the degree of $2M(M-1)n_m^2$, and number of changes will be *4N*. After performing the changes is obtained:

$$8N^3\left(\frac{M-1}{M}\right) \tag{21}$$

The necessary operations of the sum type have the same degree.

VI. RESULTS

The programming realization [7,8] of the decomposition algorithm is partially used in the development of learning plan for the discipline "*Software and internet technologies*". In technological aspect the following sequence of actions is performed:

1. The subject area is defined, as well as the purposes and relatively the tasks of the learning.
2. A scheduled plan of the specialty is developed, including determined number of disciplines.
3. For each discipline learning program is developed, including determined number of lectures.
4. The link between each pair of lectures from each pair of disciplines is established (fig. 1). The weight of the link is formed in expert way.

P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Обектно-ориентирано програмиране - 1 част (C++)	Организация на компютъра и компютърни архитектури	Синтез и анализ на алгоритми - проект	Обектно-ориентирано програмиране - 2 част (Java)	Дискретни структури	Микропроцесорна техника	Програмни системи	Графични системи	Обектно-ориентирано програмиране - 1 част - проект	Принципи на операционните системи	Програмни спецификации	Системи с бази от данни	Компютърни мрежи и Интернет	Технология на софтуерното производство	Системно програмиране	Интернет сървъри и услуги	Управление на софтуерни проекти	Програмиране за мобилни Интернет устройства	Интернет технологии	Езикови процесори	Разпределена обработка в Интернет	Проектиране на потребителски интерфейс	Извличане на информация в Интернет	Мрежово администриране
													1										
													2										
													3										
										1,2,3,4			4										
													5			1,3,4							
													6										
				9									7										
6		1,2, 3,4	6	5						3,4			8										
													9										

Figure 1. Fragment of the learning plan of the specialty and found causal links of the discipline "Technology of software production"

5. On this basis is developed the summarized graph of the specialty.
6. It is known from the carried research, that in the initial variant exist causal links in reverse direction between separate notions.
7. Partial optimization of the learning schedule according to the order of the disciplines and the volume of the included lectures is performed.

The developed algorithm and its programming realization could be used independently in the development and exploration of learning plans and programs. Modifying the limiting conditions, the decomposition allows dosing the learning units to a known degree. This makes possible the inclusion of the programming realization as an extension of the known platforms for performing distant learning.

VII. CONCLUSION

The paper presents the development of theoretically based algorithm for decomposition of finite, directed graph with weighted coefficients of the vertices and the arcs. The decomposition solution has application in determination of learning units for the purposes of distant learning. This allows development and exploration of the causal links between the disciplines in a learning plan. In practice the algorithm is partially applied in the development of learning plan for the specialty "*Software and internet technologies*".

In case of modified limiting conditions the decomposition algorithm allows adaption of the learning process to the individual abilities of the person.

References

[1] Vasileva D. D. – Adaptive software systems for e-learning, Abstract of PhD thesis, Sofia University "St. Kliment Ohridski", Sofia, 2011, 38 (in bulgarian language)

[2] Viet Anh NGUYEN, Van Cong PHAM, M. Mitev - Learner Model Based Context to Develop A Context-Aware Adaptive System in Mobile Learning. The 7th International Scientific Conference eLearning and Software for Education, Bucharest, Romania, 2011

[3] Mitev M., Racheva E., "Algorithm for initial grouping of notions in distant learning", National academy of sciences of Ukraine, Institute of cybernetics. V. M. Glushkov, International scientific journal "Controlling systems and machines", № 2, March-April 2010, pp. 13-16 (in russian language)

[4] Mitev M., E. Racheva, G. Marinova - Graph-analytic Approach for E-Learning Courses Development. Proceedings of the Fourth International Conference "Information Technologies in Education for All: Innovative Methods and Models", 2009, Kiev, Ukraine (plenary paper), UNESCO Chair "New Information Technologies in Education for All" issues, p.p. 127-136.

[5] Mitev M., E. Racheva, G. Marinova -An Algorithm for Defining of Initial Package of Notions for Purpose of E-Learning. Proceedings of the International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing, ACM International Conference Proceeding Series, ACM, USA, 2009p pp. IV.9.1-IV.9.5B, ISSBN 1313-8936, 978-1-60558-986-2.

[6] Mitev M., An. Antonov, Iv. Penev - An algorithm and program for management and evaluation of individual distance learning, 6th International Scientific Conference "eLearning and Software for Education", Bucharest, April 2010,15-16, Romania

[7] Mitev M., D. Georgiev - Program Library for Development and Research of Distance Learning Courses, ICEST – 2010, 23-26 June 2010, Macedonia, 2010

[8] Racheva E., M. Mitev et all – Project № NP 8 – Researching interdisciplinary links in specialty *"Software and internet technologies"*, Technical University of Varna, 2011