An international journal

INNOVATIVE INFOTECHNOLOGIES FOR SCIENCE, BUSINESS AND EDUCATION

Nr. 2 (3) 2008

Cover article **Mieke Vandewaetere.**
EFFECTIVE USE OF E-TOOLS IN EDUCATION: THE IMPORTANCE OF QUALITY

Hot article **Romanas Tumasonis, Inga Tumasonienė.**
DATA MINING OPPORTUNITIES IN MODERN INFORMATION SYSTEMS

VILNIUS BUSINESS COLLEGE

ISSN 2029-1035

International journal
INNOVATIVE INFOTECHNOLOGIES FOR SCIENCE, BUSINESS AND EDUCATION
is published quarterly by Vilnius Business College.
Journal (first published in 2007) is dedicated to new concepts and new insights into the issues of computer technology applications.
Short communications, review articles and full-length articles combining theoretical approaches and experimental applications are welcome.
The main emphasis is placed on the novelty, the so-called „first idea" of a new paradigm in the field of advanced programming, business realization, teaching tools etc.

TOPICS

Advanced programming
Conceptual Modelling and Databases. Natural Language Processing.
Formal Analysis and Design Methods. Algorithms and Applications. Bioinformatics

Information Technology Applications
E-Government. Grid computing systems. High-performance scientific computing
IT - business solution
Business Process Modelling, Analysis and Design. Information and Software Systems Engineering

New teaching technologies
Artificial Intelligence and Knowledge Engineering. IT in Teaching and Learning.
Computer-based Learning. Distance Education. E-tools

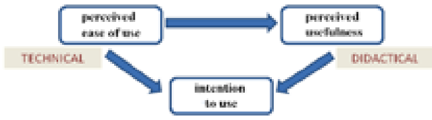
Hardware and Computer Networks
Security and Privacy

Submissions and subscription orders to be directed to:
Vilnius Business College, Kalvarijų str., 125, LT-08221, Vilnius, Lithuania,
Email: infotechnologies@kolegija.lt, http://journal.kolegija.lt

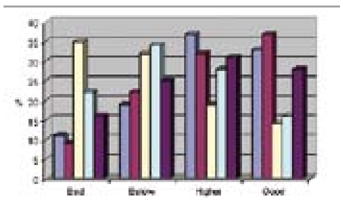
INNOVATIVE INFOTECHNOLOGIES FOR SCIENCE, BUSINESS AND EDUCATION

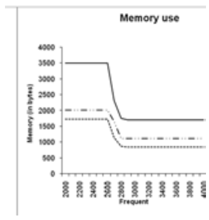
Nr. 2 (3) 2008

CONTENTS

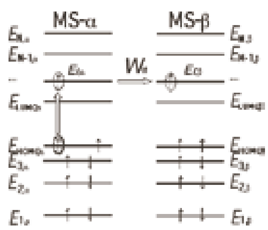


perceived ease of use
perceived usefulness
TECHNICAL
DIDACTICAL
intention to use





Memory use



MS-α
MS-β

A Word from Editors

This issue is devoted to the second part of the IIT-2007 conference proceedings. The conference took place in Vilnius, November 2007 and was a successful event which give a vise to new ideas, formed new partnerships and generally received a positive feedback.

The first section (Education) presents a thorough study of effective application of e-tools in education with emphasis on quality (Wandewaetere). Computer literacy and its problems are disclosed in short article be Lithuanian colleagues (Valavičius, Šiškevičiūtė).

The second part (Information systems) is devoted to IT application in two different fields. An algorithm for data mining sequences in bioinformatics (Tumasonis, Tumasonienė) and matrix manipulation algorithm devoted to the quantum chemical tool NUVOLA (Galikova, Gruodis) are presented and discussed.

Viltė Gridasova

EFFECTIVE USE OF E-TOOLS IN EDUCATION: THE IMPORTANCE OF QUALITY

Mieke Vandewaetere*

*Excellence Centre for Distance Learning, University College KATHO,
Dept. Teacher Training, Beernegemstraat 10, 8700 Tielt, Belgium,
ITEC, Interdisciplinary research on Technology, Education and Communication,
K.U. Leuven – Campus Kortrijk, E. Sabbelaan 53, 8500 Kortrijk, Belgium*

Received 28 August 2008, accepted 14 September 2008

Abstract. When considering e-tools, we often think too much of innovative applications- whether they can have a direct implementation in education or not. Often, e-tools are seen as novelties, people are enthusiastic about them or not, and some try applying them as a separate tool or subsequently change from an already familiar tool to a new tool. However, when we want e-tools to be durable and that applications of such tools offer possibilities in education, we need a certain kind of quality standard. A strong link between innovation and quality is inevitably needed. This study gives an overview of determining criteria for the effective use of e-tools, for both students and teachers. Additionally to this research, several suggestions are formulated for the development of a pedagogical-didactical quality measure for the use of e-tools.

Keywords. E-tools in education; quality of e-tools; didactical use of e-tools.

Short title of the paper. The didactical quality of e-tools.

**Corresponding author,*
email: mieke.vandewaetere@kuleuven-kortrijk.be, phone: +32(0)56 246111

Introduction

During the last decade, the use and implementation of e-learning or electronic learning became a fully-fledged research area and also the use of e-tools has largely intensified. More and more, learners are being prepared for living and working in the digital era [1] and e-learning in the 21st century has many faces. As a consequence, many users are overwhelmed by this technology and are rather reticent about the umpteenth new tool or application that has been launched as the ultimate application in the field of e-learning.

The area of e-learning applications has recently evolved from processes that are more focused on the distribution of information to processes where collaboration and reflection are central themes [1]. These changes in e-learning only have become possible by a continuous interaction between technical and didactical innovations in e-learning. As Ehlers [2] states, "technological applications changed from rather low innovative (traditional media) to high innovative applications, such as Web 2.0 applications. Also, pedagogical dimensions were revalued from low innovations (reaching old goals with old methods) to high innovations (reaching new goals with new methods)".

1. Computer-based learning environments

Computer-based learning environments offer learners a large spectrum of supporting applications. These applications, often imbedded in open learning environments, are developed to support learners in their learning process According to Clarebout & Elen [2], a distinction is possible between embedded software, such as feedback and the information structure in learning materials, and non-embedded software where the initiative for use depends largely on the user. Non-embedded software can be considered as added to the learning environment. This non-embedded software is often described as tools.

To have a clear wording and specification of e-tools, we can look at the categorisation suggested by Jonassen [3]. This author categorises e-tools as being part of an instructional design model for so-called constructivist learning environments. The goal of these learning environments is to encourage problem-solving and conceptual development. This occurs by confronting learners with ill-defined problems. Jonassen's system comprises the following categories.

1. Information resources. These sources offer just-in-time information which helps the learner to solve the problem. An example of this is access to the Internet.

2. Cognitive tools. These are tools that support the learner in visualising, organising, automatising or replacing thinking skills. Information maps are an example of cognitive tools.

3. Knowledge modelling tools. These tools make the understanding of the problem explicit and foster learners' reflection on their learning process. Questions like "what do I know? " and "What does it mean?" can characterise knowledge modelling tools.

4. Performance support tools. They support the cognitive functions necessary for performing a task, such as arithmetic and data-storage. By using these tools, learners can concentrate more on higher-order cognitive processes.

5. Information gathering tools. Such tools help learners in searching for certain information so that learners can stay focused on the problem solving process.

6. Conversation and collaboration tools. Social interaction is an important aspect in the learning process. Learning can be simplified by lending support to a discussion forum, a knowledge building community and a community of learners. This includes tools like e-mail, wikis, weblogs and etc.

It is obvious that the most widely known e-tools can be placed under the category of conversation and collaboration tools. Still, most electronic learning environments (ELOs) have enough possibilities to relate to other categories of e-tools, for example, the construction of a trial assessment in an electronic learning environment as support for reflecting on the learning process (tools for knowledge modeling). Other examples could be the assignment of making a mindmap (cognitive tools) or increasing information skills by integrating goal-oriented search tasks on the Internet (information gathering tools).

2. Effective use of e-tools in computer-based learning environments

Offering e-tools is one thing, however, the effective use of them is an absolutely different issue, and that is where problems often arise. A possible cause of these problems stems from the transition from pen-and-paper learning materials into web-based learning materials. The way the existing course materials are turned into web-based materials often prevents an effective and efficient use of e-tools. Most instructors, coaches or teachers simply import or drop their pen-and-paper courses into an electronic learning environment and consider the "e" of e-learning as fulfilled. However, e-learning does not only comprises electronic learning or ELO-supported learning, but also presupposes an efficient learning process. In a great number of cases people get stuck with the regular applications of e-learning, without considering the didactical approach to e-learning or the process of efficient e-learning. That is why e-learning could expand to e^2-learning or efficient e-learning.

The use of ELOs supposes a high level of learner control and the possibility that learners regulate their own learning [4]. This implies the selection of the most suitable tools to support this learning process. In creating electronic learning environments, teachers or coaches have to bear in mind several things, namely:

i) the process of tool use;
ii) the motivation for using tools;
iii) possible individual differences between learners;
iv) the efficiency and intensity that learners use the tool.

In a review of the research literature focusing on the factors influencing the use of e-tools, Clarebout and Elen [3] distinguish four possible factors. The first factor can be described as student's characteristics. According to Elen [5], several characteristics of students will influence the use of tool. However, scientific research does not yield univocal conclusions. The second factor that can influence the use of e-tools is the task and working method. Less detailed tasks will require a wider use of e-tools and tasks that give the learner a large amount of self-control will allow integrating and using more supporting tools. However, researchers have to consider the possible interaction effects between working method, learner characteristics and tool use in interpreting the inconsistent results that are frequently reported.

Explicit encouraging of using a tool is considered as a third factor influencing the effective application of e-tools. This process of encouraging includes giving advice and support, and appropriate instructions or teasers to trigger students when making assignments. Hereagain, interaction effect can have influence at this point [6]: the explicit encouraging of using e-tools is supposed only to have effect on active learners, but not on passive learners. Yet, learning style is a possibly moderating variable. Finally, the nature of the tool itself is considered as the fourth influencing factor. Of all tool categories that are comprised in Jonassen's above-mentioned system, the performance support tools (arithmetic tools and data-storage) and information gathering tools have been most frequently used by students [7]. Nevertheless, only the purpose of the tool and the type of support that the tool offers is considered. More inherent characteristics such as the tool quality have not been discussed in literature, as far as it is known. As Ehlers [8] states, we have to bear in mind that the quality of a learning process cannot be captured in terms of something that is offered via a learning environment to the learner, which is similar to a unidirectional approach of e-learning. Quality of a learning process can rather be described as a process of coproduction and interaction between the learner and the learning environment. This is a more multidirectional approach to e-learning. To put it differently, it is the learner who defines to what extent the learning environment can be considered as successful and in what way the e-tools will be effectively used. The provider of the learning process, the teacher, trainer or coach, however, can assure that a maximally suitable and high-qualitative learning environment will be created in which the chance for success is optimized.

Ehlers [9] states that, according to learners, the quality concept comprises more than only instructional of technological interface design. Based on a large-scaled survey, Ehlers was able to distinguish seven quality fields, briefly described in table 1[9].

Table 1. Ehlers' quality fields

Quality field	Description
Instructor support	Includes a two-way interaction, in which the learner not only receives feedback but also gives feedback to the instructor. This field also includes active support of the learning process and individual support.
Collaboration	Consists of active creation of knowledge, group activities, possibilities for online discussion, etc.
Technology	Adaptability and personalisation are the core words. As well, learner must have the possibility for synchronous communication (e.g. chat, videoconferencing) and consulting the e-course materials off-line.
Costs and benefits	This is closely related to the expectations learners have. The higher the expectations, the higher the benefits and the lower the costs should be. The costs and efforts that are expected from learners have to be in subjective proportion to the benefits that are related to the course. However, this depends on learners to a high extent.
Information transparency	The main questions at this point are as follows: do learners get advice before they start the course? Who is the helpdesk in case of technical problems? What is the main goal of the course? An overview of the course also improves the quality.
Course structure	Providing a manual or workshop to introduce the learning environment heightens the quality of the learning process. Also, providing intervening tests and/or assessments via the electronic learning environment is a positive factor.
Didactics	This category often overlaps with the previously described categories and comprises the following features: extending the learning environment with access to background materials, presentation materials, goal-oriented structuring of the course, the fact that the e-course cannot be considered as an appendix of a pen-and-paper course, feedback on the learning process by offering assessments and tests, and the availability of individualised tasks.

Other studies [9; 10] also distinguish several dimensions that are determining the success of electronic learning environments, such as the information quality, system quality, quality of service, system use, user satisfaction and net profit.

The elements above can clearly be in aid of a needs analysis. Needs often are classified in two categories: needs referring to the content and needs referring to the organization. Often, these needs analyses are made from the e-tools developers' point of view. As a result, there is a difference between theoretical quality of e-tools and the experienced quality. Table 2 gives an overview.

Table 2. Types of quality

Factor	Theoretical quality	Practical quality
evaluator	developers	users
influence of context	context independent	context dependent
evaluation	cognitive	cognitive/affective
attitudes	positive or no influence	positive/negative
beliefs	positive or no influence	positive/negative

The quality of e-tools and of e-learning environments as evaluated by the users (instructors and learners) is largely dependent on the context and goals. Besides cognitive factors, there are affective factors that play a role in the evaluation of tools (e.g., motivation, openness for new things). The frequently undervalued influence of attitudes and beliefs is often not considered when analyzing the quality of e-learning.

When we want e-learning to be revalued to e^2-learning (efficient electronic learning), we have to bear in mind two additional issues. First, there are the pedagogical-didactical needs, or questioning what an instructor wants to reach by means of the electronic offering of learning materials. If the answer on this question can only be formulated from the point of the organization (e.g. the teachers have to start applying e-learning; there is no other option as the license has been bought etc.), then e-learning has probably little chance to succeed. However, if an instructor is eager to support or enhance the learning process by using e-tools, then this motivation increases the chance of an effective and successful use of the applications, both for the instructor and the learners. A second group of needs are technological by nature. Technological foreknowledge and requirements and an eventual lack of such are one of the main thresholds preventing users to start using e-learning of building up an ELO.

The central question in the below-presented research is what kind of requirements - didactical and technological- an e-tool has to comply with in order to realize a positive attitude of users that leads to an effective use of the tool. This question is set both with instructors and with learners. The following sections describe the method of data collection, the results and conclusions. Further research is discussed and the last section offers an overview comprising guidelines and possibilities for qualitative implementation of e-tools in course design.

3. Methodology

3.1. Participants. All students and lecturers of the University College KATHO (Belgium) were considered as eligible for the research, since all of them have experience with on-line learning environment (based on Blackboard®) that is used in education offered by KATHO. KATHO has seven departments, offering 18 practice-oriented basis trainings, with 34 specialisations (bachelor degree). The students and lecturers came from all departments. The questionnaire was placed in the KATHO on-line learning environment. In a period of 10 days, 25 lecturers and 316 students answered the questionnaire.

3.2. Materials. The questionnaire was divided into two parts. In the first part, technical aspects that may influence the effective use of e-tools were listed. The second part gave an overview of didactical aspects. Participants were asked to give their personal top three in both lists. They had to indicate what they considered as the three most important technical and didactical criteria that e-tools should meet to be effectively used. Participants also had the possibility to add new criteria.

3.3. Procedure. All students and lecturers received an e-mail with the aim of the research and a link to the questionnaire. In order to bring about a univocal interpretation of the e-tool concept, a short description was given before the start of the questionnaire. Participants had to fill in the questionnaire at once and were prompted when they left one or more items blank.

4. Results

4.1. Technical aspects. According to the lecturers, there are three most important technical aspects, namely: in the first place, the tool has to be simple in use (68% of all respondents); second, it has to be time-saving (52% of the respondents); third, it has to reduce the amount of paper (36% of the respondents). The students pointed out the following factors: simplicity of the tool (55% of the respondents) stands first; the fact that a tool has to be free of advertising (44% of the respondents) was chosen as second, and thirdly, the tool has to work fast (39% of the respondents). The eye-catching fact found by the research is that both the students and lecturers are willing to work with several applications, and only few respondents (16% of the lecturers and 12% of the students) emphasized that it is vitally necessary that the tool can be integrated in other programs or learning environments. The language in which the tool is offered (Dutch interface) is important for only 1/4 of the lecturers (24%) and students (23%). Extensions of possibilities to personalise tools are highly valued by tool developers, but less by the users: only 8% of the lecturers and 10% of the students consider personalisation as a condition for effective tool use. Table 3 gives the values (percent of users that placed the criterion in their top three) for each technical criterion.

Table 3. Technical criteria and importance according to lecturers and students

Technical aspects	Lecturers, (%)	Students, (%)
offline use	0	10,71
simplicity	68,00	55,09
no high memory required	4,00	16,71
fast download	0	20,89
integration in other software	16,00	12,27
free of adverts	32,00	43,60
Dutch interface	24,00	22,72
fast	28,00	39,43
simple navigation	28,00	16,97
time saving	52,00	33,94
less paper	36,00	25,85
no fast connection required	8,00	4,70
personalisation/extensions	8,00	10,18
other	0	0,26

The additional criteria given by students were mainly individual and are summarized in Table 4. From the results it is obvious that a "good" tool has to be simple in use in the first place.

Table 4. Technical criteria – individual additional comments

Technical aspects. Individual comments by users
- One can choose his/her own user name and password.
- The tool has to be clear, especially for people using the tool for the first time.
- Memory of use, e.g., task search, auto-fill etc.
- Content can be copied to other programs (e.g. text editing programs).
- The tool does not take up too much space/memory.
- The tool must work without problems. Sufficient information has to be provided, e.g. a conveniently arranged manual.
- The tool is easily accessible.
- An amusing interface is always good.
- The tool may not cause conflicts with other software.
- The tool is reliable (no crashes).
- The tool is always available.
- The tool has to be universal and can be available for other schools and institutions.
- The tool can be used on other platforms (e.g., Linux).

4.2. Didactical aspects. According to the lecturers, in the top three of didactical aspects the following two criteria share the first place: a tool has to foster interactivity and has to lead to a more efficient collaboration (both aspects have won 44% of the respondents' replies). For the lecturers, a tool also has to lead to more simplified communication (40% of the respondents). The students place a more efficient collaboration on the first place (51%), followed by more simple communication (48%) and on the third place is the flexible supply of learning materials (47%). Obviously, both groups consider the support of communication and the flexibility of the learning process as the most important criteria when evaluating e-tools. The aspect whether a tool was successfully used by peers (lecturers or students) previously in other situations or courses (good practice) or not has little influence on the use of an e-tool (lecturers - 8%; students- 11%). The attitude of other lecturers or other students also has little impact on the effective use of a tool (lecturers- 20%; students- 12%). Using a tool only because it will enhance the course or learning materials to a higher level is important only for 1/4 of the lecturers (28%) and one fifth of the students (18%). Again, this fact emphasizes the necessity for a pedagogical-didactical justification of the implementation of e-tools – it cannot be limited to mere using the tool just because such a tool exists. Table 5 gives the values (percent of users that placed the criterion in their top three) for each pedagogical-didactical criterion.

Table 5. Pedagogical-didactical criteria and importance according to lecturers and students

Pedagogical-didactical aspects	lecturers, (%)	students, (%)
more simple communication	40,00	47,52
flexible learning materials	36,00	47,26
more simple assessment	12,00	14,36
more interactive	44,00	37,60
requires no training	28,00	34,47
higher level of working	28,00	18,02
course exceeding didactical goals	24,00	20,10
easy transformation from pp to e-content	12,00	10,71
successful use in other courses	8,00	11,23
positive attitude of users	20,00	11,75
more efficient collaboration	44,00	50,65
other	0	0,78

The additional didactical criteria that were provided were mostly content-related and are represented in Table 6. It is apparent that from a pedagogical-didactical perspective the flexibility of communication and the learning process are crucial. The main function of e-tools is supportive.

Table 6. Pedagogical-didactical criteria – individual additional comments

Didactical aspects. Individual comments by users
- "Doing new things in new ways" (Marc Prensky).
- On the one hand, the relation between instructor and learner is more intense (interactive), on the other hand, one can develop a more clearly structured self-study package.
- A tool has to establish good communication between the provider and the receiver.
- Tools help to understand the course better. They give you the feeling that you are not alone and that support is always available. They also give the possibility to participate actively in the learning process.
- The content of a tool has to be up-to-date; it cannot comprise any outdated information.
- The tool may not be laborious, but has to be effective in use.
- Not only exercises and solution can be provided, but also courses and handbooks can be available (e.g., pdf, e-book)
- School and instructors has to use the tool logically and consistently, so that the learners do not have to look at several places if they have missed anything.
- Subjective opinions are less possible.
- Thanks to e-tools, one can work always/everywhere. You are not limited to a particular time.
- The tool supports the course materials.
- A tool may not be obliging, only supportive for those who really want to use it.

5. Conclusions

This study takes a first step toward a more interactive collaboration between developers or providers of e-tools and electronic learning environments on the one hand and the users of these applications on the other hand. Lecturers and students of the University College KATHO were interviewed about what they considered as most important criteria in order to effectively use an e-tool. Criteria were divided into technical and didactical criteria. According to lecturers, a good e-tool is easy to use, is time-saving and helps to reduce the amount of paper. Also, e-tools should support an interactive learning process and lead to a more efficient and simple communication. According to students, an e-tool has to be easy to use, has to work fast and has to be free of advertising. Moreover, encouraging an efficient collaboration and simple communication is a necessity to function as a good tool. Learners also expect that a good tool makes it possible to offer the learning materials in a flexible way (*anytime anywhere* content).

The main conclusion of this research is that as far as the technical aspects are concerned, the learners and instructors have low expectations and demands. Features as software compatibility, integration in other software, interface design, etc. are considered as nice to have, but not need to have. They are not considered as necessary in order to qualify the e-tool as a good tool. Closely interrelated with this is the KISS-rule: "keep it short and simple" or "keep it stupidly simple". When considering the pedagogical-didactical aspects, communication-related criteria take a prominent place. Tools serve for supporting and enhancing the communication and have to make the learning process more flexible. Related to the didactical criteria expectations are rather high, and exactly these criteria are often considered as less important when implementing e-tools.

6. Discussion discussions

What determines whether e-learning in an organisation or institution can become successful or not? This question is at the basis of many debates and research concerning the quality of e-learning. Still, the quality concept is often looked at unilaterally, from the view of the developers, and needs analyses often are performed from an organisational point of view. Questions that are posited refer to the content that will be offered (what?) and the method of offering (how?). In both questions and their answers, the content is often subservient to the tools that will be used. In many situations, a tool is selected first, after which the possibilities of the tool are explored and instructors search how they can incorporate the tool as much as possible. This can be considered as a top-down approach: from the tool to the content. A more valuable approach, resulting in a greater chance to effective e-learning is the bottom-up approach. With this approach, instructors ask how they can improve their course materials, how they can enhance the quality of the learning process by modifying their course. Based on the didactical needs for the course modification, instructors can select which tools suit the best in order to reach their goals. This bottom-up approach starts from the didactical point of view en puts the content and didactical methods first.

There should be an interaction between the possibilities and features of e-tools and the content. If the learning materials are of that kind that it is more efficient to offer them in pen and paper, there is no sense in trying to fit in these materials in an electronic learning environment. Based on the technological and pedagogical-didactical needs of users, e-tools have to be further developed, optimised and adapted, and, vice versa, based on more refined e-tools, pedagogical-didactical methods can be evaluated, adjusted or completed. The important and necessary issue is a continuous interaction between technical innovations and pedagogical-didactical innovations. This calls for a new way of thinking for both the tool developers and instructors. E-tools developers should think about the needs of instructors or learners and ask themselves which learning-related goals can be reached by using the tool. It could be as well useful if instructors and learners think more like developers and ask how the tool can be improved. Alternatively, they can participate in discussions on software improvement. By doing so, users and instructors learn how to reflect critically on their use of tools, can interact with other users that are more or less advanced in using the tool and have the possibility for a broader and more intense use of e-tools. Such discussions may be useful for tool developers as well. By having a direct contact with the base (i.e. the users), tool developers can promptly improve e-tools. Based on users' real-life positive or negative experiences with e-tools or some aspects and parts of those, it is easier to adapt e-tools in a user-oriented way.

Electronic learning environments and e-tools are becoming more oriented towards the learner and ask for active feedback by the users of the learning environments. As a consequence, there is strong need for a pedagogical-didactical support when implementing e-tools and when transforming pen-and-paper courses into electronic learning materials. The use of a tool because such a tool exists is not enough and there is a need for a structured approach when one wants to start using e-learning of implementing e-tools in a highly qualitative and well-considered way. Related to this, we can refer to the Technology Acceptance Model (TAM [11]). This model states that the intention of effective use of a tool largely depends on the evaluation or appraisal of the tool use. Usability can be related to the technical quality criteria, the usefulness of the tool is comparable with the didactical criteria. Fig. 1 shows a completion of the TAM-model, adjusted for the quality concept.

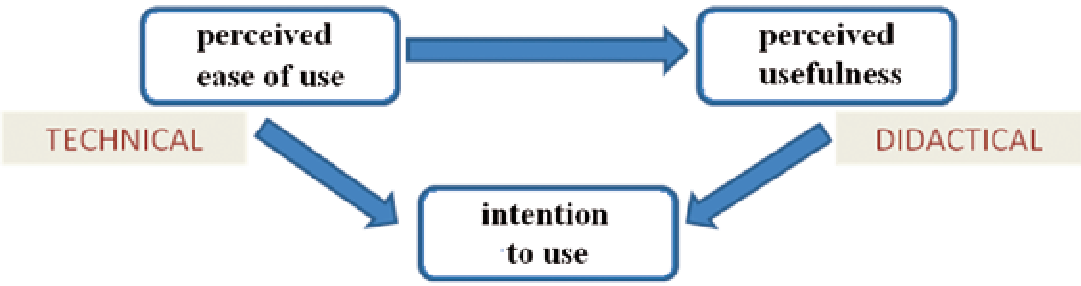


Fig. 1. Completion of the Technology Acceptance Model, adapted for quality.

It is worth paying attention to the gap between what an electronic learning environment can offer and the expectations that the users have to this environment in terms of didactical and technical criteria. Indeed, recent research shows that there is still a large discrepancy between desires and wishes of the users and what technology can offer [12].

Based on the definition of Ehlers' quality construct [1], the following section contains the material consists of suggestions (specific for higher education with electronic learning environment) to improve the quality of the e-learning process, based on a pedagogical-didactical point of view. Obviously, not all quality fields will be equally represented because some applications are still in the stage of development, testing or evaluation. The overview below (see Tables 7 to 14) should not be considered as a standard or fixed approach, but rather as providing guidelines with possibilities to reflect when evaluating the quality of electronic learning materials. Also, this overview of practical implementations offers a window on the broad range of possibilities that exist to go further than simply import pen-and-paper materials into an electronic learning environment.

Table 7. Practical implementation of theoretical quality constructs – Support provided by the instructor

Quality field 1: Support by the instructor	
Possibility for two-way interaction	
WHAT?	Learners do not only receive feedback but also send feedback to the instructor of the course.
HOW?	By means of intermediate assessments of tests, questioning the course content. This can help the instructor in the further refinement and improvement of the learning materials for following learners.
Moderation of the learning process	
WHAT?	Active moderation of the learning process in a communicative way.
HOW?	Discussion forum in the ELO with a forum "questions to the instructor".
Learner-oriented interaction versus content-oriented interaction	
WHAT?	The learner-oriented interaction puts more emphasis on the personal learning process of the learner, while the content-oriented interaction focuses on the communication between the instructor and the learner.
HOW?	Both types of interaction can be obtained via discussion fora. By asking directed questions (by the instructor) the learners will be able to start a reflection process.
Individualised learner support	
WHAT?	Focuses on the specific needs of the learners and foresees additional information according to the interests of the learner(s).
HOW?	A page with links, additional documents or organising theme-based videoconferences with active participation of the learners.

Table 8. Practical implementation of theoretical quality constructs – Collaboration and communication in the course

Quality field 2: Collaboration and communication in the course	
Social collaboration	
WHAT?	An online course should focus on collaboration with other learners, experts, mentors or instructors.
HOW?	On-line group discussions (chat, videoconferencing, mail-newsgroups etc.).
Discursive collaboration	
WHAT?	Collaboration where the emphasis is not on the social aspect, but rather on the actively creating knowledge.
HOW?	Group work via weblogs, creating wiki's, websites, making movies, developing a manual (also possible with the pen-and-paper approach).

Table 9. Practical implementation of theoretical quality constructs – Technology

Quality field 3: Technology	
Adaptability and personalization	
WHAT?	Can the learning environment be adapted according to the personal preferences of the user? Also providing the facility that after logging in the users can start where they logged out the last time.
HOW?	Most electronic environments offer a broad range of personalisation facilities (e.g., colors, lay-out, adding figures etc.). Both for the instructor and for the learner.
Possibilities for synchronous communication	
WHAT?	Does the learning environment have a synchronous communication facility?
HOW?	Use references to communication tools (chat tools) in the links-page of the electronic learning environment.
Availability of content	
WHAT?	The learning material should be available in several formats (preferably in pdf-format). The learners should have the possibility to download the materials on their own computer and consult the materials offline.
HOW?	Protected text editing documents of creating documents in portable document format (pdf). Large files can be zipped in maps.

Table 10. Practical implementation of theoretical quality constructs – Costs, expectations, benefits

Quality field 4: Costs, expectations and benefits	
Expectation of individualization	
WHAT?	The expectations that the learners have regarding the flexibility of online learning and its individualised aspects, concerning the content as well as the support and guidance.
HOW?	Permanent announcement on the electronic learning environment gives the learner an overview of what he/she can expect. Make sure that the learners' expectations fit closely to the content that an instructor can and wants to offer (downsize if necessary).
Individual non-economic costs	
WHAT?	The efforts that are required from the learners to stay motivated and concentrated for a course with a strong individualised route.
HOW?	Posting announcements on the electronic learning environment, on a regularly basis, encouraging e-mails to all learners halfway the course, before an exam etc.
Economic costs	
WHAT?	Financial costs.
HOW?	A link to the Internet providers that offer cheap(er) rates for instructors and learners.
Practical advantages	
WHAT?	Learners expect practical advantages of online learning in their daily lives.
HOW?	A page with links to interesting websites, a weblog with applications of the course content, stories of the graduated learners, best practices etc.
Interest in the use of the course and media	
WHAT?	Learners are not only interested in the content but also in the e-tools that support the learning process.
HOW?	Maximum integration of e-tools in the tasks, assignments, group work etc. A diverse stock of best practices based on the feedback from the previous learners, the previous experiences.

Table 11. Practical implementation of theoretical quality constructs – Information transparency

Quality field 5: Information transparency	
Support and advice	
WHAT?	Important dimension of subjective evaluation of the quality is the type of support that the learners receive before they start working in the electronic learning environment.
HOW?	Providing a demo session with the introduction of the course, where the most important functionalities of several tools are discussed. It is better to make the learners learning by doing instead of giving them a paper manual.
Information concerning the organization	
WHAT?	Learners consider it important to have background information and especially to know where they can find particular parts of the course, e.g., training forms, exam regulations, contact information of the co-learners and representatives, links to relevant websites, the course schedule etc. This information can be provided in the electronic learning environment via e-tools. It has to be clearly presented where and how the learner can access this information.
HOW?	A search function on the general (home) website, a search function for the course website, separate intranet for the learners, structured outline of the materials (per course, per year, per target group, per month etc.)
Information concerning the goals and content	
WHAT?	Learners prefer detailed information for each course that will be offered electronically.
HOW?	Regular update of the study information documents, course documents, ECTS-files has to be available in the electronic learning environment or on the website.

Table 12. Practical implementation of theoretical quality constructs – Course structure

Quality field 6: Course structure	
Personal support of the learning process	
WHAT?	The importance of personalised and individualised course support.
HOW?	Via additional contents, extending an adaptive learning route or referring to other information sources. Being available by e-mail or by synchronous communication tools (e.g., live chat sessions between the instructors and the learners).
Introduction to technical aspects and content	
WHAT?	This is strongly related to "support and advise" in Quality field 5: information transparency. Here, the content is also discussed.
HOW?	At the beginning of an e-learning course, a face-to-face meeting can be organised for the instructor and the learners. At this meeting, the learners can be split into groups, additional questions can be asked and a demonstration of course content can be offered.
Assessment	
WHAT?	The possibility of online assessment.
HOW?	By means of interim assessments, the instructors can demonstrate the type of questions that learners can expect at the examination. The learners receive (electronic) feedback for the outcome. Alternatively, the score for the interim online assessment can add up for the general summative score.

Table 13. Practical implementation of theoretical quality constructs – Didactics

Quality field 7: Didactics. Part 1	
Background materials	
WHAT?	The importance of access to the background materials. This aspect has already been discussed more than once in the above-mentioned quality fields. This fact emphasizes particular importance that the learners expect from this feature.
HOW?	A page with links to other information sources, stimulation information processing skills by integrating search tasks in the assignments, organising a theme evening or a videoconference with the experts, a live chat session with the experts with the possibility of asking questions to the experts, instructors or learners.
Multimedia enriched presentation materials	
WHAT?	Many learners express their need for the learning materials that are enriched with a broad range of multimedia.
HOW?	Integration of audio, movies, authentic texts etc. in the learning materials. The reference to the multimedia enriched materials on every page of the course or in every learning object does not yield the desired effect. It is better to evaluate the existing assignments and the content to what extent these materials can be offered differently. For example, describing a movie fragment is also possible by creating a link to a video-sharing channel where that movie or fragment is available. Also, the description of a sound can be described equally well by integrating an audio file in the learning environment.
Structured and goal-oriented course materials	
WHAT?	Arranging the course materials according to the goals that the instructor wants to achieve with that course, and not arranging the materials according by the content of chapters.
HOW?	When constructing a new course, an overview of the goals that have to and may be reached can be created (basic goals and extended goals). On the basis of this overview, a course can be built up. As soon as the course materials are available, each chapter/part may comprise an overview of the goals that are targeted at in this particular chapter/ part. This is also possible at the end of each chapter/ part where the learners get an overview of what they are expected to know/learn/be able to do/etc. after studying that part.

Table 14. Practical implementation of theoretical quality constructs – Didactics

Quality field 7: Didactics. Part 2	
Support of (life long) learning	
WHAT?	Working in and with an electronic learning environment must learners allow to acquire specific learning skills and become more experienced in their competencies for lifelong learning.
HOW?	Stimulating information processing skills by teaching the learners how to search efficiently, how to deal with the large amount of available information, teach them how they can process and summarise this information (e.g., by making mindmaps, concept cards or diagrams).
Feedback on learning progress	
WHAT?	Integration of interim assignments in the learning environment as support and feedback on the learning process.
HOW?	The stipulation here is that the learners also receive feedback for the scores of the assignments. Rather than giving a rough score (e.g., a number), the trainer can choose to provide qualitative feedback with personal comments. Another possibility is to let the learners make a short reflection task following every assignment they have to hand in. The question for such a reflection task could be, for example: what do you like in this assignment? what are your strong and weak points in this assignment? how could you lower or increase the score etc. In this way, the learners often gain new insights into their own learning styles. Such insights are necessary in order to revise the assignment and adjust it before handing in.
Individualised tasks	
WHAT?	Tasks that are developed to meet the needs of individual learners.
HOW?	The development of individualised learning routes. By screening the results or by providing substantial feedback, the instructor can position learners on the scale and give them appropriate tasks. The differentiation in the electronic learning environment is simpler that integrating differentiation in a classroom-situation, since e-learning itself is already a more individualised process. Starting with similar individualised learning routes and the development of additional supporting materials (for all learners) requires a great effort from the instructor.

Acknowledgements
The author would like to thank the Department of Teacher Training, KATHO PHO Tielt.

References

[1] Ehlers, U.-D. (2007). *Competencies for Achieving High Quality in E-Learning.* Presentation at ICL 2007 Conference, Villach (Austria).

[2] Clarebout, G., & Elen, J. (2006). Tool use in computer-based learning environments: towards a research framework. *Computers in Human Behavior, 22,* 389-411.

[3] Jonassen, D. H. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Ed.), *Instructional-design theories and models. A new paradigm of Instructional Theory* (Vol. II, pp. 215-239). Mahwah, NJ: Lawrence Erlbaum.

[4] Hannafin, M. (1995). Open-ended learning environments: Foundations, assumptions, and implications for automated design. In R. D. Tennyson & A. E. Barron (Eds.), *Automating instructional design: Computer-based development and delivery tools* (pp. 101-129). Berlin: Springer.

[5] Elen, J. (1995). *Blocks on the road to instructional design prescriptions.* Leuven: University Press.

[6] Lee, Y. B., & Lehman, J. D. (1993). Instructional cueing in hypermedia: a study with active and passive learners. *Journal of Educational Multimedia and Hypermedia, 2*, 25-37.

[7] Oliver, K. M., & Hannafin, M. J. (2000). Student management of web-based hypermedia resources during open-ended problem solving. *The Journal of Educational Research, 94,* 75-92.

[8] Ehlers, U.-D. (2004). Quality in e-learning from a learner's perspective. *European Journal of Open, Distance and E-Learning,* 2004/I.

[9] Wang, Y.-S., Wang, H-Y., & Shee, D. Y. (2007). Measuring e-learning systems success in an organizational context: Scale development and validation. *Computers in Human Behavior, 23,* 1792-1808.

[10] DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of Management Information Systems, 19(4),* 9-30.

[11] Saadé, R., & Bahli, B. (2005). The impact of cognitive absorption on perceived usefulness and perceived ease of use in on-line learning: An extension of the technology acceptance model. *Information & Management, 42(2),* 261-386.

[12] Cardoso, V., & Bidarra, J. (2007). Open and Distance Learning: Does IT (Still) Matter? *European Journal of Open, Distance and E-Learning,* 2007/I.

COMPUTER LITERACY SURVEY

Eugenijus Valavičius*, Daiva Šiškevičiūtė

Vilnius Business College, Kalvarijų street 125, Vilnius, Lithuania

Received 14 January 2008, accepted 17 June 2008

Abstract. This paper analyses and compares the results of two surveys on computer literacy. The questionnaire survey was divided into three parts, namely: computer and user interface, information management, text-hypertext-multimedia. The second survey was aimed: to spot and analyze changes in students' computer literacy, compare computer literacy level of students studying at universities and those studying at higher education colleges, suggest possible improvement to IT as a subject taught at higher education institutions and to establish computer literacy skills development after students have participated at IT classes at the college.

Keywords: computer literacy, information technologies, survey of IT knowledge and skills.

Short title of the paper. Computer literacy.

*Corresponding author, email: eugvalav@vpu.lt

Introduction

This paper analyses and compares the results of 2 surveys on computer literacy. The first survey was run in Vilnius Pedagogical University during 2001-2002 (coordinated by E. Valavičius). The participants of this survey were university students (following the studies in humanities) and secondary school students. The questionnaire survey was divided into 3 parts, namely: computer and user interface, information management, text-hypertext-multimedia [1].

The second survey was aimed at the following goals: to spot and analyze changes in students' computer literacy, compare computer literacy level of students studying at universities and those studying at higher education colleges, suggest possible improvement to IT as a subject taught at higher education institutions and to establish computer literacy skills development after students have participated at IT classes at the college. This survey uses the same questionnaire as mentioned above.

Vilnius Business College students were acting as the participants of the second survey, however, the authors are targeting at extending this survey to a number of colleges and universities.

1. Results of the first survey

The participants of the first survey were divided into 4 groups depending on time since last of IT classes they had attended: i) first-year students (86 persons); ii) second-year and higher (94); iii) part-time students (37); iv) secondary school students (58). According to the survey results, the participants of each group were distributed into 4 subgroups - 2 groups with computer literacy above average and 2 groups- below average.

The questions were divided into three topics:

i) computer (hardware, user interfaces, basic concepts);
ii) information management;
iii) text (also hypertext and multimedia).

The questionnaire comprised no questions to evaluate motor skills due to the fact permanent knowledge (which level was aimed to establish) is better described by basic concepts and definitions. Here are some examples of questions that the questionnaire comprised:

i) label the main computer parts;
ii) find definitions of a folder, file, document, font, paragraph, hypertext, multimedia etc;
iii) find links among several enumerated terms.

The evident leaders for first two parts of questions were secondary school students while university students were leaders answering the third part of the questionnaire focusing on text, hypertext and multimedia. A closer look at the results shows that secondary school students were best in the basic concepts subtest (about 70% fall into 2 higher groups) and students of the first course were best in the information management subtest (about 80% of students are above the average).

Fig. 1 represents the results of the third subtest focusing on text, multimedia and hypertext.

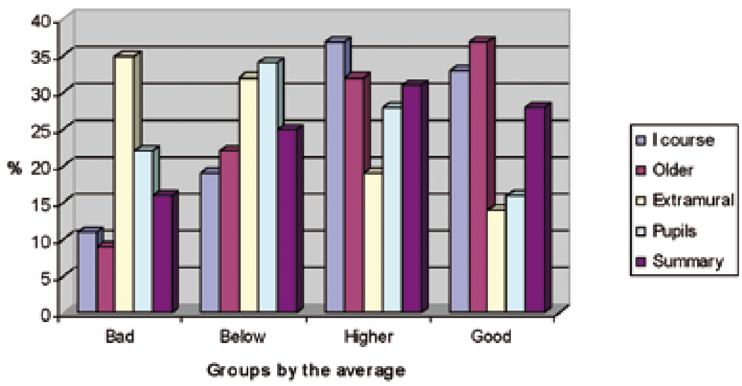


Fig. 1 Results of the subtest on text and hypertext.

Fig. 1 shows that text definitions are easiest to understand for the first year and older full-time students but not part-time students. Secondary school students did not succeed in this subtest: only about 40% of then scored higher than the average. However, the total average scores (31% as higher and 27% as good) in this subtest show that the knowledge level of text operations is better compared to the other two subtests: less than 40% in 2 higher groups in the computer subtest and near 50% in the information management subtest [2].

The conclusions of this survey can be following: operations with text, hypertext and multimedia are easy topics of computer literacy (the summary column- the last column in each of 4 groups - shows that nearly 60% of the students scored above average), however, this particular field of computer literacy is underestimated at secondary schools. The names of elements of the user interface and common definitions are easy to forget: only 40% of respondents knew them.

2. Results of the second survey

The secondary school curriculum of information technologies has undergone significant changes since 2005. Computer hardware and networks in schools have improved as well. Consequently, college and university teachers have to understand and accept these changes so that to find IT topics that are most relevant for their students.

The participants of the second survey (carried out in the autumn of 2007) were 173 first-year students studying at Vilnius Business College.

The maximum value of each question was 1 point. The average scores of subtests are shown in Fig. 2.

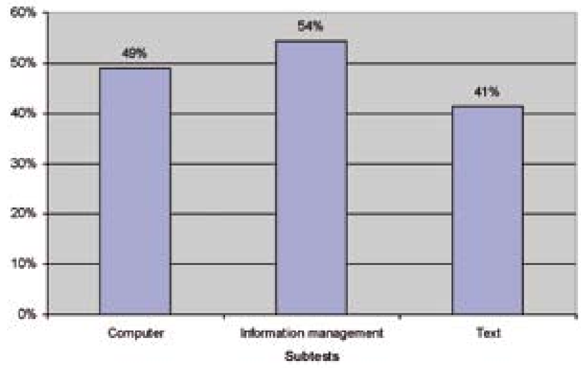


Fig. 2. Averages of subtests

We can see the best average score has been reached in the second subtest focusing on information management; the scores on the computer and user interface issues were less than 50% and the lowest average score is in the third subtest. However, these average scores do not reflect the real knowledge of the students because of the unequal distribution of very high and very low scores in individual tests.

Fig. 3 shows the distribution of the number of students whose scores are below and higher than the average of each subtest.

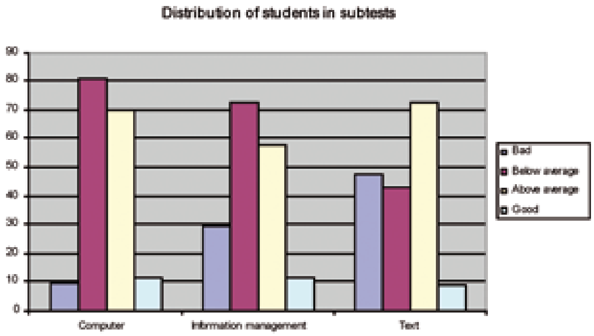


Fig. 3. Distribution of students' scores in subtests.

This diagram represents the level of knowledge of the field in each subtest. Students' scores are distributed into four groups: „Bad", „Below average", „Above average", „Good".

The biggest number of students fall into the „Below average" group in the first and second subtests while the biggest group is „Above average" in the third subtest. However, the third subtest has the lowest number of students that belongs to the group who has scored as „Good". This is the main reason why the average score of the third subtest was the lowest among all subtests (41%, as Fig.2 shows). Another reason for that can be that in this subtest the highest number of students fall into group who scored as „Bad".

When we look at the distribution, it is easy to notice that more than 60% of students scored below average in the second subtest. The scores of the other two subtests have almost equally distributed below and above the average. However, the average score is highest in the second subtest because of the total big number of highest scores.

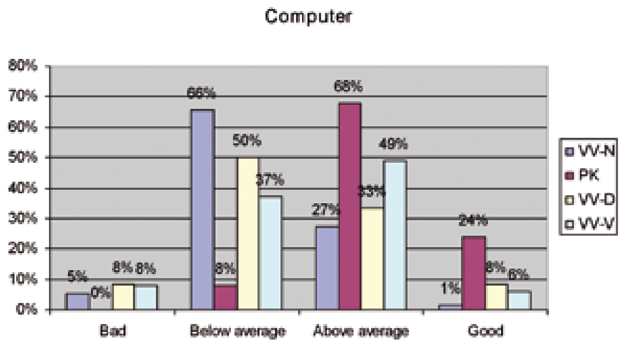


Fig. 4. The distribution of knowledge related to the computer according to study programs and study lines.

In Fig. 4, we see the level of knowledge in the first group of questions related to computer issues. It is easy to see that full-time students of the study programme of Computer Programming (PK-D) are evident leaders. None of them has scored under the category "Bad". Besides, they scored most in the categories "Above average" and "Good". Students of the study programme of Business Administration following evening studies (VV-V) scored about average: 37 % of them are "Below average" and 49 % - "Above average", whereas 8 % scored as "Bad" and 6 % as "Good". The scores of the full-time (day-time) students of Business Administration are similar to the evening students, with a slight shift to lower scores. If we ask why Computer Programming students have showed high results, the answer is obvious: they come across computer-related definitions, elements of user interface and other IT-related terms much more frequently compared to the students of other study programmes. The day-time Computer Programming students have showed lower scores. This can be explained by vast work experience and higher motivation of the evening class students who are mostly working specialists; some of them are older than daytime students. Nowadays, as most workplaces are often equipped with computers, it has positive influence in the knowledge of IT.

On the other hand, the results showed by the part-time students are significantly lower than those of day and evening classes. The majority of the part-time students who participated in the survey are much older than others, as they graduated from secondary school 5, 8, 10 or more years ago. They do not use IT at all or use it very rarely. Another factor is that many part-time students have family which means that they can spend less time for learning. However, more than 1/5 of them scored above average.

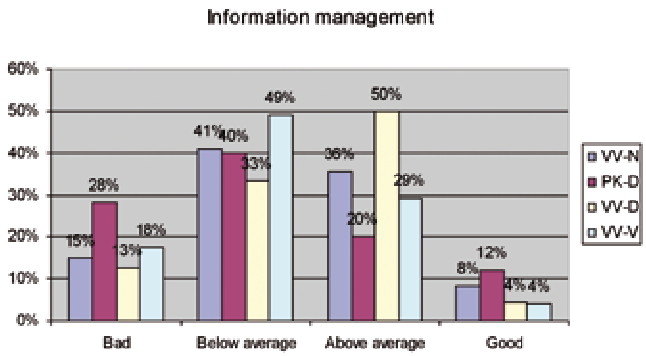


Fig. 5. The distribution of scores of knowledge related to the information management according to study programmes and study lines.

Computer Programming students (PK-D) are leaders in the group of questions related to information management. Nevertheless, they are leaders in the lowest scores as well (40% scored below average and 28 % scored as bad). These results are even worse than those of the part-time students. It seems that this group of students is low-motivated and they are chosen to study in order to obtain a diploma but not acquire knowledge. At the same time, more than a half of the students of Business Administration (daytime class) fall into the categories "Above average" and "Good". It means that totally the knowledge of the students of Business Administration is best.

If we compare these results with results of the other two groups of topics, we can notice that the questions related to information management are most difficult. This is the only topic where total scores are shifted to the left (low or below average) side. Our teaching experience shows that usually most of students know how to create a directory, copy, move or delete files, which are the main operations of file system of computer. However, the basic (theoretical) definitions are not so easy to understand and remember.

Looking at the scores of the topic related to Text, we can see that answers of the bigger part of the students fall under the categories of "Above average" and "Good". For example, 64% (56% + 8%) of the full-time Computer Programming students and 54% (50% + 4%) of the full-time Business Administration students (daytime class) are acquainted with this field better than the average. The evening Business Administration students are an exception – most of them scored as "Good". Nevertheless, at the same time most of them (35%) have showed low scores.

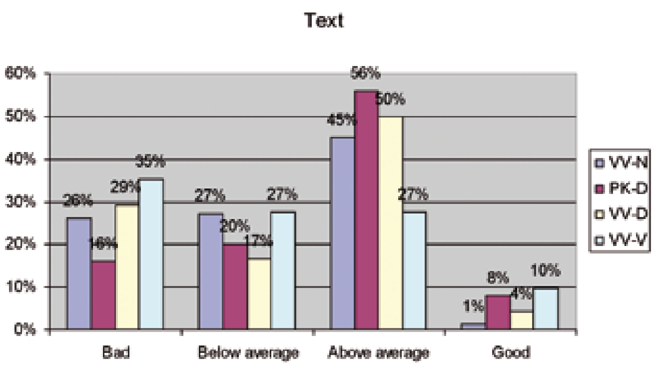


Fig. 6. The distribution of scores of knowledge related to the text processing (including hypertext and multimedia) according to study programmes and study lines.

The issue of text processing is important not only for IT specialists. This is a necessary part of knowledge for all educated people. Let us have a closer look at this field. The test on this topic consists of questions which are not only related to text as it is but also includes the definitions related to multimedia" and hypertext. While working with computer, we use the word text when we see various types of information (text, pictures, animations etc.) composed in one window. On the other hand, when there is an attempt to understand on a more thorough level, one has to know definitions and principles of using each type of information. In some cases, people merely do not want to think about it, sometimes they do not know the Lithuanian equivalents (e.g., many of us use the word multimedia in Lithuanian instead of its Lithuanian equivalent daugialypė aplinka). That is why we left only questions related to text directly ("Pure" text) and excluded multimedia and hypertext issues when revising the survey results.

Fig.7 shows the average score of the knowledge related to text. We see that almost all average scores reach the 50% level. It is very high level if we compare with other questions or groups of questions. The best scores were achieved by the full-time students of Business Administration (daytime studies) – about 55%. Computer Programming students appeared not to score best in this test. We have tried to explain this fact because of their specific interests – they are interested mostly in specific IT fields requiring profound knowledge and less in operations that regular users perform. As an unexpected exception in this question we can mention the high score achieved by part-time students of Business Administration, on the one hand, and the surprisingly low results showed by the evening Business Administration students.

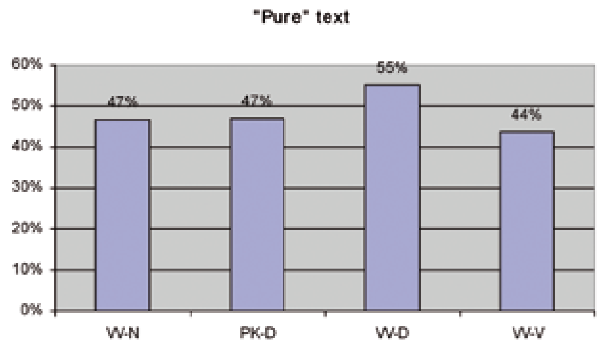


Fig. 7. The distribution of scores of knowledge related to the text according to study programmes and study lines.

However, the Computer Programming students scored best in the tests focusing more complicated definitions of IT terms. Table 1 show the parameters of total average knowledge related to multimedia and hypertext. As the maximum value of each question is 0.5, the average knowledge level is 0.25.

Table 1. Parameters of total average knowledge related to multimedia and hypertext

Study Programme	Multimedia	Hypertext
Business Administration (part-time)	0.14	0.15
Computer Programming (full-time, day-time studies)	0.22	0.26
Business Administration (full-time, day-time studies)	0.08	0.17
Business Administration (full-time, evening-time studies)	0.09	0.16

We see that the Computer Programming students have scored around average in the field of multimedia issues and above average in the field of hypertext. It means that 1 of 2 students understand these topics. The knowledge of students of other study programmes are about 1/3 and below. The issues of hypertext are better-known whereas the definition of multimedia is much (even 2 times) less known.

3. Preliminary comparison of the results of the two surveys

One of the purposes of the second survey was to compare the knowledge that students bring to a higher education institution after the secondary school. We have made an attempt to make a hypothesis that knowledge is improving every year. However, when we compare the preliminary results of the survey, we can state that there are no significant changes concerning the students' knowledge, which students had 5 years ago and which they bring from secondary schools now. Of course, we have to emphasize that our tests included only the main concepts and definitions and no motor skills.

In both surveys, the biggest part of students scored as „Below average", but there were bigger number of students that scored as „Bad" 5 years ago than now in the first subtest. We can state that now students are more accustomed to the computer and user interface.

The absolute leaders in the second subtest (information management) were secondary school students five years ago. However, we got unexpected results in the second survey: the first-year students in Vilnius Business College know this topic much below average. We could find only one reason for such results: the group of pupils that took part in the first survey was not randomly, and, thus, properly chosen. They were participants of the courses organized for secondary school students by Vilnius Pedagogical University, which means that their knowledge level knowledge was higher.

We obtained better good results in the third subtest (related to text and hypertext) several years ago. On the other hand, we have to emphasize that the participants of the first survey took this questionnaire after the IT course training at university while now we have analyzed the results before teaching IT discipline at the college. Apart from that, there can be another reason too: the usage of computer now is more oriented towards the Internet and entertainment rather than to the process of creation.

Conclusion

These two surveys are separated by a long period of time and significant changes in IT teaching programs in secondary schools, nevertheless, the preliminary comparison shows similar results. The correlation is evident when comparing youngest participants of the first survey and the first year students. In both cases, the answers about concepts and definitions of text and hypertext scored 45% above average and in both cases the concepts of computer architecture and Windows interface are easier to understand than the basics of text and hypertext. However, we cannot present the full comparison because our survey participants have not accomplished the IT course at the college.

The preliminary analysis of the present survey allows us to presuppose that college IT teachers can spend less time teaching computer interface in favor of information management and other topics.

References

[1]. E. Valavičius, Būsimųjų mokytojų IT žinių tyrimas (in lith.) // Kompiuterininkų dienos – 2003. V.: Žara, 2003, p. 263-267.

[2]. ECDL program, 4.0 version. [http://www.ecdl.lt/modules/tinycontent/index.php?id=2] (retrieved 2007.10.15)

DATA MINING OPPORTUNITIES IN MODERN INFORMATION SYSTEMS

***Romanas Tumasonis*[1,2]**, Inga Tumasoniene*[3]

[1]Vilnius College, Jasinskio street 15, Vilnius, Lithuania
[2]Vilnius Business College, Kalvariju street 125, Vilnius, Lithuania,
[3]Vilnius Gediminas Technical University, Sauletekio al. 11, Vilnius, Lithuania

Received 12 February 2008, accepted 17 June 2008

Abstract. The paper deals with the search and analysis of the subsequences in large volume sequences (texts, DNA sequences etc.). A new algorithm ProMFS for mining frequent sequences is proposed and investigated. The algorithm is based on the estimated probabilistic-statistical characteristics of the appearance of the elements of the sequence and their order. The algorithm builds a new much shorter sequence and makes decisions on the main sequence in accordance with the results of analysis of the shorter one.

Keywords. Data mining, text mining, frequent sets, frequent sequences.

Short name: Data mining opportunities

** Corresponding author, email: r.tumasonis@eif.viko.lt*

Introduction

The task of association mining is to discover a set of attributes shared among a large number of objects in a given database. For example, consider the books database where the attributes represent authors or books and the objects represent customers. The discovered patterns are the set of books most frequently bought together by the customers. An example could be that 35% of the people who buy Den Brown's *The Da Vinci code* also buy *Angels and Demons*. The store can use this knowledge for promotions, shelf placement and etc. There are many potential application areas for association rule technology, which include, store layout, catalogue design [1], customer segmentation [2], telecommunication alarm diagnosis and so on.

The general task of discovering all frequent associations and rules in very large databases is quite challenging. The search space is exponential in the number of database attributes, and with many of database objects the problem of input/output minimization becomes paramount. Though, most current approaches are iterative in nature, requiring multiple database scans, which is obviously very expensive. Some of the methods, especially those using some form of sampling, can be sensitive to the data-skew, which can adversely affect performance [3]. Also, most approaches use very complicated internal data structures which have poor locality and add additional space and computation overheads. Our goal is to overcome all of these limitations [4].

Many algorithms combining the features listed above depend on the database format, the decomposition technique, and the search procedure used. Most popular algorithms is Eclat [5] (Equivalence CLAss Transformation), MaxEclat [6], Clique [7], MaxClique [8], Top-Down [9], and AprClique [10]. Our algorithms not only minimize input/output costs by making only one database scan, but also minimize computation costs by using efficient search schemes. The algorithms are particularly effective when the discovered frequent item sets are long. Our tide list-based approach is also insensitive to the data-skew.

1. Data mining algorithms

The task of discovering all frequent sequences in large databases is quite challenging. The search space is extremely large. For example, with *m* attributes there are *O(m k)* potentially frequent sequences of length *k*. With many millions of objects in the database, the problem of input/output minimization becomes paramount. Most algorithms are iterative in nature and requiring as many full database scans as the longest frequent sequence, which is naturally very expensive.

Assume that we have a set *L*, consisting of *m* distinct elements, also called *items*.

$$L = \{i_1, i_2, ..., i_m\} \tag{1}$$

An itemset is a nonempty unordered collection of items. A sequence is an ordered list of itemsets. A sequence ($\alpha_1 \to \alpha_2 \to ... \to \alpha_n$) is denoted as α, where the sequence element α_j is an itemset. An item can occur only once in an itemset, but it can occur multiple times in different item sets of a sequence. We solve a partial problem, where itemset consists of one item only. A sequence α is a subsequence of another sequence β:

$$\alpha = (\alpha_1 \to \alpha_2 \to \alpha ... \to \alpha_n), \tag{2}$$

$$\beta = (\beta_1 \to \beta_2 \to ... \to \beta_n), \tag{3}$$

if there exist such numbers $t_1, t_2, ..., t_n$, where $t_{j+1} = t_j + 1$, $j = 1, ..., n$ and $\alpha_j = \beta_{t_j}$ for all a_j. Here, β_{t_j} are elements of the set *L*. We analyze the sequence (the main sequence) *S* that is formed from single elements of *L* (not their sets, as in the classical formulation of the problem). In general, the number of elements in *S* is much larger than that in *L*. We have to find the most frequent subsequences in *S*. The problem is to find subsequences whose appearance frequency is more than some threshold called *minimum support*, i.e. the subsequence is frequent if it occurs in the main sequence no less frequently than the minimum support [11].

The main problem is as follows: it is necessary to define only potentially useful subsequences, check up and prolong them gradually. Two algorithms are analyzed that analytically eliminate such subsequences that actually cannot be frequent. Two implementations of *Generated-Sequence-Pattern algorithm* (GSP) with and without economizing memory are tested. A recursive approach to sequence mining is suggested. It enables saving memory, too. The algorithms are compared and the cases, in which one algorithm works better than another one, are determined,

The new algorithm for mining frequent sequences ProMFS is based on the estimation of the statistical characteristics of the main sequence: i) the probability of an element in the sequence; ii) the probability for one element to appear after another one; iii) the average distance between different elements of the sequence

The main idea of the algorithm is the following:

i) some characteristics of the position and interposition of elements are determined in the main sequence;

ii) much shorter new model sequence $\tilde{C}$ is generated according to these characteristics;

iii) new sequence is analyzed with the GSP algorithm (or any similar one);

iv) frequency of subsequences in the main sequence is estimated by the results of the GSP algorithm applied on the new sequence.

Let us assume the following set *L=acc. to Eq. (1)* as the set consisting of *m* distinct elements. Probability of occurrence of element i_j in the main sequence where $i_j \in L$, $j = 1, ..., m$ could be calculated as follows:

$$P(i_j) = \frac{V(i_j)}{VS}, \tag{4}$$

V(i$_j$) represents the number of elements i_j in the main sequence *S*; *VS* represents the length of the sequence. Note that

$$\sum_{j=1}^{m} P(i_j) = 1, \tag{5}$$

Probability $P(i_j \mid i_v)$ represents the appearance of element i_v after element i_j, where $i_j, i_v \in L$, $j, v = 1, ..., m$. Note that

$$\sum_{v=1}^{m} P(i_j \mid i_v) = 1, \tag{6}$$

for all *j*=1 ... m.

Distance $D(i_j \mid i_v)$ between elements i_j and i_v, where $i_j, i_v \in L$, $j, v = 1,\dots, m$, represents the number of elements that are between i_j and the first found i_v from i_j to the end of the main sequence (i_v is included). The distance between two adjacent elements of the sequence is equal to 1. Set of average distances represents a matrix of average distances Ã. Elements of the matrix are as follows:

$$a_{jv} = \text{Average}\,(D(i_j \mid i_v),\ i_j, i_v \in L),\ j, v = 1,\dots, m. \tag{7}$$

All these characteristics can be obtained during one search through the main sequence. According to these characteristics, a much shorter model sequence $\tilde{C}$ is generated, the length of which is *l*. Denote its elements by c_r, $r = 1,\dots, l$. The model sequence $\tilde{C}$ will contain elements from *L*: $i_j \in L$, $j = 1,\dots, m$. For the elements c_r, a numeric characteristic $Q(i_j, c_r)$, $r = 1,\dots, l$, $j = 1,\dots, m$, is defined. Initially, $Q(i_j, c_r)$ is the matrix with zero values that are specified after the statistical analysis of the main sequence. The complementary function $\rho(c_r, a_{rj})$ is introduced that increases the value of characteristics $Q(i_j, c_r)$ by 1. The first element c_1 of the model sequence $\tilde{C}$ is that from *L* corresponding to $\max(P(i_j))$, $i_j \in L$. According to c_1, the function

$$\rho(c_1, a_{1j}) \Rightarrow Q(i_j, 1 + a_{1j}) = Q(i_j, 1 + a_{1j}) + 1 \tag{8}$$

$j = 1,\dots, m$, is activated. Remaining elements c_r, $r = 2,\dots, l$, are chosen in the way described below. Consider the *r*-th element c_r of the model sequence $\tilde{C}$. The decision, which symbol from *L* should be chosen as c_r, will be made after calculating $\max(Q(i_j, c_r))$, $i_j \in L$. If for some p and t we obtain that

$$Q(i_p, c_r) = Q(i_t, c_r) \tag{9}$$

then element c_r is chosen by maximal value of conditional probabilities, i.e. by

$$\max(P(c_{(r-1)} \mid i_p), P(c_{(r-1)} \mid i_t)) \tag{10}$$

$c_r = i_p$ if following condition is satisfied:

$$P(c_{(r-1)} \mid i_p) > P(c_{(r-1)} \mid i_t) \tag{11}$$

and $c_r = i_t$ if following condition is satisfied:

$$P(c_{(r-1)} \mid i_p) < P(c_{(r-1)} \mid i_t) \tag{12}$$

If these values are equal, i.e.

$$P(c_{(r-1)} \mid i_p) = P(c_{(r-1)} \mid i_t) \tag{13}$$

then c_r is chosen depending on $\max(P(i_p), P(\mid i_t))$. After choosing the value of c_r, the function

$$\rho(c_r, a_{rj}) \Rightarrow Q(i_j, r + a_{rj}) = Q(i_j, r + a_{rj}) + 1 \tag{14}$$

is activated. All these actions are performed consecutively for every $r = 2, \dots, l$. In this way, we get the model sequence that is much shorter than the main one and that may be analyzed by the GSP algorithm with much less computational efforts.

We have changed one characteristic (the average distance between different elements of the sequence) to the frequent distance between different elements of the sequence. $\tilde{F}$ is the matrix of these frequent distances. The elements of the matrix are described follows:

$$f_{jv} = \text{Frequent}\,(D(i_j \mid i_v),\ i_j, i_v \in L),\ j, v = 1,\dots, m., \tag{15}$$

2. Results of the research

Let us compare these two different implementations and recursive algorithm by a special example. Suppose we have a text file with 90 000 A and B symbols. Our goal is to find all the frequent sequences. We have compared the time expenditure and memory use. The results are shown in Fig. 1 and Fig. 2

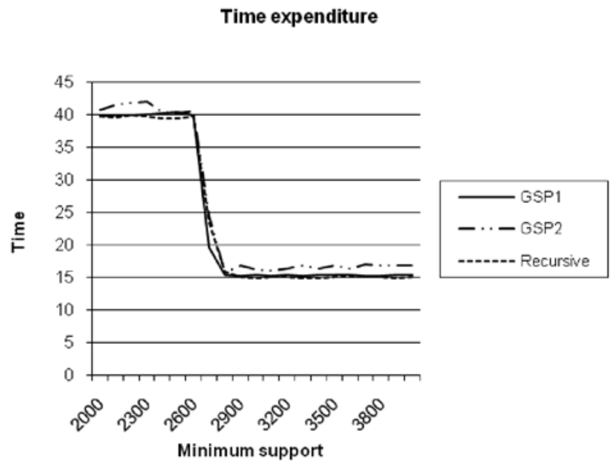


Fig. 1. Time expenditure of GSP implementations and recursive algorithm.

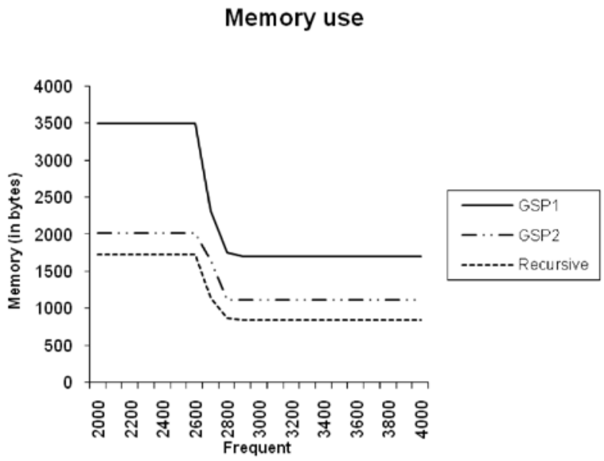


Fig. 2. Memory use of GSP implementations and recursive algorithm

The probabilistic mining of frequent sequences was compared with the GSP algorithm. We have generated a text file of 100,000 letters (1000 lines and 100 symbols in one line). L={A, B, C}, i.e. m=3, $i_1 = A$, $i_2 = B$, $i_3 = C$. In this text we have included one very frequent sequence ABBC. This sequence is repeated 20 times in one line. The remaining 20 symbols of the line are selected at random. First of all, the main sequence (100,000 symbols) was tested with the GSP algorithm. The results are presented in Fig. 3 and Fig. 4. They will be discussed in more detail together with the results of ProMFS. ProMFS generated the model sequence $\tilde{C}$ of length l=40.

This model sequence was examined with the GSP algorithm using the following minimum supports: 8, 9, 10, 11, 12, 13, and 14. The results are presented in Fig. 3 and 4. Fig. 3 shows the number of frequent sequences found by both GSP and ProMFS. Fig. 4 illustrates the expenditure of computation time used by both GSP and ProMFS to obtain the results of Fig. 3 (the minimum support in ProMFS is Ms=8; the results are similar for larger Ms). The results in Fig. 3 indicate that if the minimum support in GSP analyzing the main sequence is comparatively small (less than 1500 with the examined data set), GSP finds much more frequent sequences than ProMFS. When the minimum support in GSP increases from 2500 to 6000, the number of frequent sequences by GSP decreases and the number of those by ProMFS increases. In the range of [2500, 6000], the number of frequent sequences found by both GSP and ProMFS is rather similar. As the minimum support in GSP continues growing, the number of frequent sequences found by both algorithms becomes identical. When comparing the computing time of both algorithms (see Fig. 5), we can conclude that ProMFS operates considerably faster. In the range of the minimum support in GSP [2500, 6000] ProMFS needs approximately 20 times less of computation time as compared with GSP to obtain the similar result.

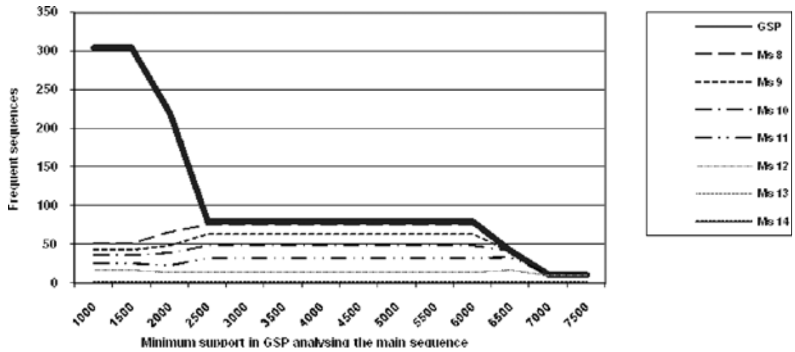


Fig. 3. The number of frequent sequences found by GSP and ProMFS (the minimum support in ProMFS is Ms=8, … ,14).

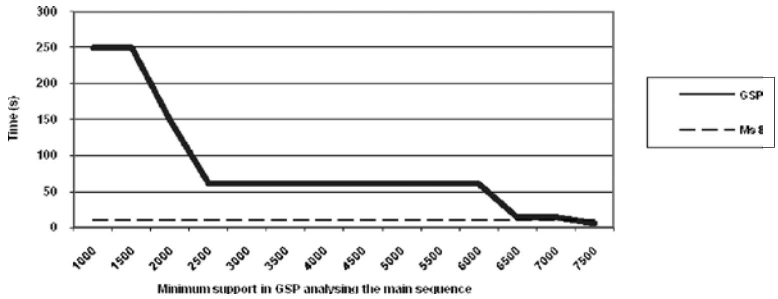


Fig. 4. The computing time by GSP and ProMFS (the minimum support in ProMFS is Ms=8).

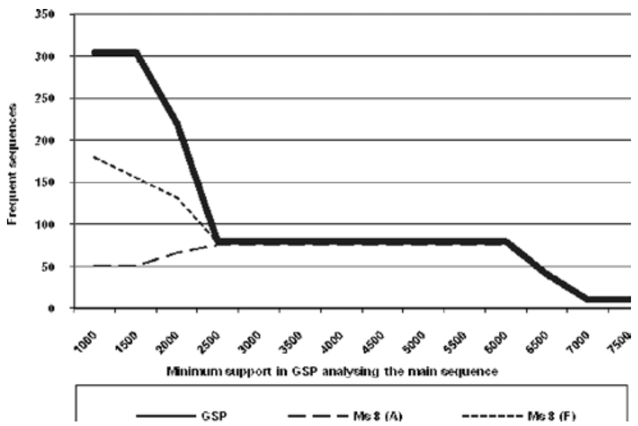


Fig. 5. The computing time by GSP and ProMFS (the minimum support in ProMFS is Ms=8).

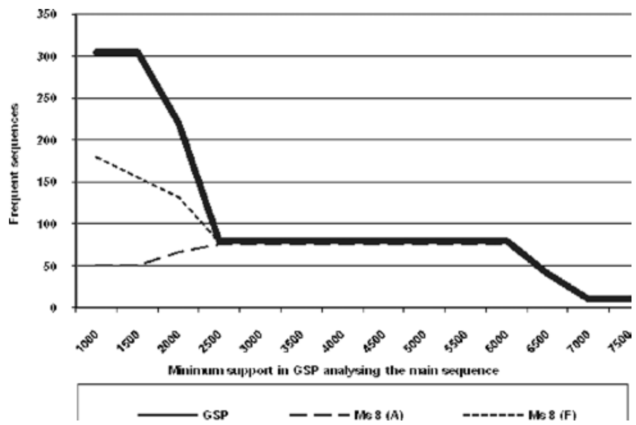


Fig. 6. The number of frequent sequences found by GSP and ProMFS with the average matrix (Ms 8 (A)) and the frequent matrix (Ms 8 (F)) (the support in ProMFS is Ms=8).

Fig. 6 illustrates the expenditure of computation time by both GSP and ProMFS with two different matrixes to obtain the results of Fig. 3 (the minimum support in ProMFS is *Ms*=8; the results are similar for larger *Ms*). The results in Fig. 3 indicate that if the minimum support in GSP analyzing the main sequence is comparatively small (less than 1500 with the examined data set), GSP finds many more frequent sequences than ProMFS. But the results of ProMFS with the matrix of frequent distance are better than with the matrix of average distance.

Databases are designed for programs to process automatically; text is written for people to read. Text mining methods can be used in bioinformatics for analysis of DNA sequences. *A genetic DNA sequence* is a succession of special symbols (letters) representing the primary structure of a hypothetical or real DNA strand or molecule. The four letters of sequence are A, C, G, and T, representing the adenine, cytosine, guanine, thymine, which is nucleotide subunits of a DNA. The DNA sequence printed with no spaces between letters: AAACTGGTCTGAC [2].

We have worked with real DNA data from [12] (*34565 Homo sapiens* chromosome 1 genomic contig). The size of the database is equal to 250 MB. The results are shown in Fig. 7 and Fig. 8.

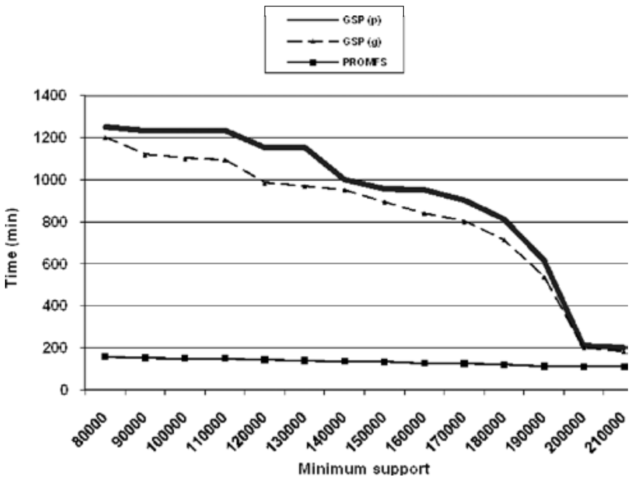


Fig. 7. The computation time by both GSP and ProMFS in *Homo sapiens* chromosome 1 genomic contig.

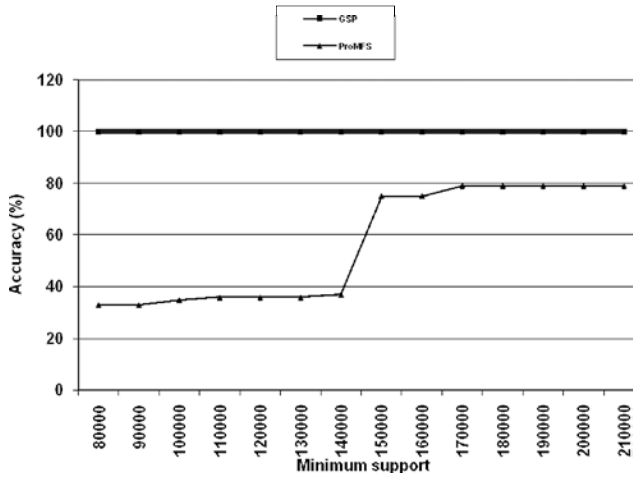


Fig. 8. The accuracy by both GSP and ProMFS in *Homo sapiens* chromosome 1 genomic contig.

Conclusions

Two implementations of the Generated sequence pattern (GSP) algorithm and a recursive algorithm have been examined. The first implementation disregards saving memory, while the second one minimizes the memory consumption. Both implementations and a recursive algorithm are time-intensive. Therefore, intense memory use requires relatively little time and is observed in the case of large data sets. The best memory saving algorithm is recursive algorithm because this algorithm uses memory just for frequent sequences, but not for all generated candidates.

The new algorithm ProMFS for mining frequent sequences with matrix of frequent distance has been proposed. It is based on the estimated probabilistic-statistical characteristics of the appearance of elements of the sequence and their order: the probability of an element in the sequence, the probability for one element to appear after another one, and the frequent distance between different elements of the sequence. The algorithm builds a much shorter new model sequence and makes the decision on the main sequence in accordance to the results of analysis of the shorter one. The model sequence may be analyzed by the GSP or other algorithm for mining frequent sequences: the frequency of subsequences in the main sequence is estimated by the results of the model sequence analysis.

The experimental research indicates that the new algorithm modification enables saving the computation time to a large extent and loses fewer sequences compared with the older algorithm modification, which is especially important when analyzing very large data sequences.

References

[1] Agrawal R.C., Agrawal C.C., Prasad V.V. (2000) Depth first generation of long patterns. *In Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Boston, Massachusetts. P. 108-118.

[2] [http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html], retrieved 09/09/2007

[3] Zaki M.J. (2000) *Parallel sequence mining on shared-memory machines.* In: Zaki, M.J., Ching-Tien Ho (eds): Large-scale Parallel Data Mining. Lecture Notes in Artificial Intelligence, Vol. 1759. Springer-Verlag Berlin Heidelberg New York. P. 161-189.

[4] Zaki M.J. (2001) *SPADE: An efficient algorithm for mining frequent sequences.* Machine Learning Journal, 42 (1/2) p. 31-60 (Fisher, D. (ed.): Special issue on Unsupervised Learning).

[5] Pei P.J., Han J., Wang W. (2002) Mining Sequential Patterns with Constraints in Large Databases. *In Proceedings of the 11th ACM International Conference on Information and Knowledge Management (CIKM'02),* McLean, VA. P. 18-25.

[6] Pinto P., Han J., Pei J., Wang K., Chen Q., Dayal U. (2001) Multi-Dimensional Sequential Pattern Mining. *In Proceedings of the 10th ACM International Conference on Information and Knowledge Management* (CIKM'01), Atlanta, Georgia. P. 81-88.

[7] Pei J., Han J., Mortazavi-Asl B., Pinto H., Chen Q., Dayal U., Hsu M.-C. (2001) PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth. *In Proc. 17th International Conference on Data Engineering ICDE. Heidelberg.* P. 215-224.

[8] Han J., Pei J. (2000) FreeSpan: Frequent pattern-projected sequential pattern mining. *In Proc. Knowledge Discovery and Data Mining.* P. 355-359.

[9] Ayres J., Flannick. J., Gehrke J., Yiu T.(2002) Sequential pattern mining using a bitmap representation. In Proc. *Knowledge Discovery and Data Mining.* P. 29-435.

[10] Zaki M.J. (2000) Scalable algorithms for association mining..*Knowledge and Data Engineering* IEEE Transactions onVolume 12, Issue 3.

[11] Kum H.C., Pei J., Wang W. (2003) ApproxMAP: Approximate Mining of Consensus Sequential Patterns. *In Proceedings of the 2003 SIAM International Conference on Data Mining (SIAM DM '03)*, San Francisco, CA. P. 311-315.

[12] [ftp://ftp.ncbi.nih.gov/genbank/], retrieved 09/09/2007.

NUVOLA: INTERMOLECULAR ELECTRON TRANSFER PROPERTIES USING QUANTUM CHEMISTRY APPROACH

Nurija Galikova*, Alytis Gruodis

Department of General Physics and Spectroscopy, University of Vilnius, Saulėtekio al. 9, building 3, LT-10222 Vilnius, Lithuania

Received 14 January 2008, accepted 19 June 2008

Abstract. The project NUVOLA represents a new quantum chemistry tool based on the perturbation theory methods. The simulation of the dynamical processes between molecular aggregates using quantum chemistry approach is discussed and described. Using the Born-Oppenheimer assumption, the intermolecular electron transfer (IET) process was modeled based on the Fermi golden rule. The estimation of IET rate between two molecular systems could be allowed in frame of electronic charge redistribution between atomic orbitals, molecular orbitals, separate atoms, molecular fragments, donor and acceptor pairs. Quantitative IET processes can be analyzed in framework of energies as well as spatial charge redistribution.

Keywords: NUVOLA, intermolecular electron transfer, Fermi golden rule, Born-Oppenheimer assumption, one-particle approach.

Short title of the paper. NUVOLA: intermolecular electron transfer.

* *Corresponding author, e-mail: Nurija.Galikova@ff.vu.lt*

Introduction

Theoretical interpretation of the electronic properties of nanostructures and molecular clusters could be performed empirically (partially by laws of classical mechanics, low successful way) as well as semi-empirically or *ab-initio* (by means of several quantum mechanics approaches, high successful way). During the last decade, different simulations of molecular properties have allowed estimating the probable molecular structure and ground electronic state properties: molecular charge distribution, vibration frequencies, multipole/dipole/quadrupole moments etc. [1].

Our attempts are devoted to simulate the dynamical physical processes using quantum chemistry approach. Intermolecular electron transfer (IET) processes play an important role in the energy exchange between ordered molecular aggregates. Numerical publications dating back to the 1960s are devoted to the description the phenomena in general [2, 3]. An IET process can be presented in two following approaches:

i) empirical mechanistic model (spatial shift of electron as particle);
ii) wave interaction model (coupling of corresponding waves representing electron behaviour).

The mechanistic description of an IET process which expresses the movement of an electron between two separated molecular systems (MS) during redox reactions could be expanded two electron exchange formalism [4, 5]. An IET process can be named as the fundamental in understanding basic principles of complicated biological and biochemical reactions [6]. A quantum chemistry approach as a semi empirical or ab-initio approach could be titled as the universal approach for modelling the IET reactions independently from the object origin [7, 8].

This work is devoted to present the project NUVOLA [9] – a quantum chemistry tool based on the perturbation theory methods – and to outline the most relevant mathematical problems as well as presumable solutions of IET rate computing.

1. Fermi Golden rule

Let us assume the existence of two separated molecular systems MS-1 and MS-2 as presented in the Fig.1. IET reactions between MS-1 and MS-2 are estimated in the framework of the electronic charge redistribution between two charge-donor and charge-acceptor molecules according to the Mulliken model.

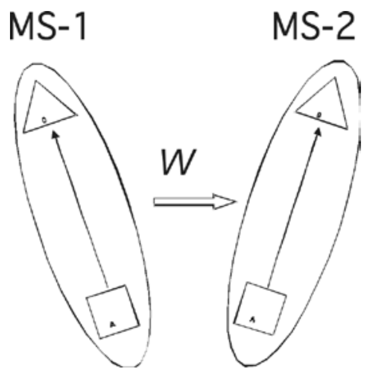


Fig.1. Schematic representation of the IET process (ratio *W*) between two molecular systems MS-1 and MS−2. Molecular system consist conditionally of charge donor D and charge acceptor A parts.

We will use the Born-Oppenheimer assumption, so the wave-functions α and β representing the behaviour of the systems MS-1 and MS-2, respectively, could be described using two separated factorized parts: electronic and nuclear. The calculation of IET rate is based on the Markus theory [10] and the well-known equation called Fermi Golden rule [11], which expresses the resonant process between two electronic states (state energy E_α and E_β, respectively) of the space-separated objects:

$$W_{\alpha\beta} = \frac{2\pi}{\hbar} S^2_{\alpha\beta}\, \delta\left(E_\alpha - E_\beta\right). \qquad (1)$$

Intermolecular electronic coupling $S_{\alpha\beta}$ could be presented using the Dirac notation as interaction of two wave-functions $_\alpha$ and $_\beta$, belonging to the MS-1 and MS-2, respectively.

$$S_{\alpha\beta} = \left\langle \alpha \middle| V_{\alpha\beta} \middle| \beta \right\rangle, \qquad (2)$$

where $V_{\alpha\beta}$ represents the matrix element of intermolecular interactions between such two MS. δ-function plays the key role in order to determine the strong resonant condition:

$$\int_{-\infty}^{+\infty} \delta(x)\, dx = 1 \,. \qquad (3)$$

Total transition rate W will be expressed using Eq.(4):

$$W = \frac{2\pi}{\hbar} \sum_{\alpha,\beta} \left\langle \alpha \middle| V_{\alpha\beta} \middle| \beta \right\rangle^2 \delta\left(E_\alpha - E_\beta\right) \,. \qquad (4)$$

Fig. 2 represents the energetic scheme of the IET process between two electronic states (labeled *initial* (i) and *final* (f)), where the state energy is denoted as $E_{i\alpha}$ and $E_{j\beta}$. These two states belong to separated MS – MS-1α and MS-2β, respectively. Further, we will operate in the terms of molecular electrons. Each electronic state is restricted by the Pauli principle, which allows only two molecular electrons with opposite spins in one state. The mechanism of IET supposes electron excitation from an occupied to a suitable unoccupied molecular orbital (MO), and then IET is permissible in case of an electronic overlap between two electronic states of molecules.

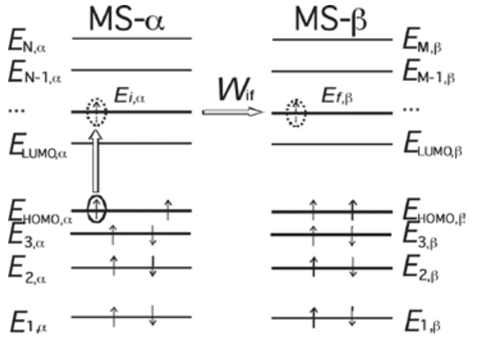


Fig. 2. Energetic scheme of the intermolecular ET process (ratio W_{12}) between two electronic states $E_{i\alpha}$ and $E_{j\beta}$, belonging to the separated molecular system MS-1α and MS−2β, respectively.

A simulated process could be described as follows:
i) one-electron excitation in the MS-1, from HOMO and lower state to LUMO and higher state;
ii) deactivation of excited state in backward-way (this process is not important in this case);
iii) repeatable interaction between corresponding electronic states of different MS (transition rate W_{if}).

Probability of IET (or transfer rate) is linearly proportional to the square of electronic overlap coefficient of corresponding molecular orbitals. As a result, the highest probability peaks of the IET spectrum belong to the states with equal energy called *resonant states*.
In the cases where the presented equation is not convenient for routine calculation, we offer a noticeably simplified form:

$$W_{\mathrm{if}} = \frac{2\pi}{\hbar} E_{\mathrm{if}}^2 S_{\mathrm{if}}^2 \delta\left(E_{\mathrm{i}} - E_{\mathrm{f}}\right), \qquad (5)$$

where Eif represents the excited state energy, S_{if} – dimensionless coupling between two electronic states. We have used a dumped resonant condition using the well-known Gauss functions through the selected value σ which represents the half width at half-maximum (HWHM):

$$\delta\left(E_{\mathrm{i}} - E_{\mathrm{f}}\right) \Rightarrow \frac{1}{2\pi\sigma} \exp\left(- \frac{\left(E_{\mathrm{i}} - E_{\mathrm{f}}\right)^2}{2\sigma^2} \right). \qquad (6)$$

The detailed description of the calculation routine and derivation of Eq. (5) are presented in Ref. [12]. The advanced IET process simulation between ordered molecular aggregates is presented in Ref. [13].

2. Hardware and software

NUVOLA can be run on a variety of UNIX operating systems. It supports the *OS LINUX FEDORA 4* on LINUX workstation (machine-independent 32 or 64 version). The programs are written in *C++*. All floating-point computations are performed in precision of 64-bits. Standard compiler *gcc* was used without additional libraries.

3. Realization of algorithm

The project *NUVOLA* is a quantum chemistry tool for the IET rate calculation, which handles *Gaussian'03* [14] program package as a contributory tool. Several sets of molecular orbitals (MO) of investigating molecules are required, which could be taken from output files of the mentioned programs. The requested data (the spatial positions of atoms of both MS) is transformed into three files.

NUVOLA contains three operational components called *links*. *C*++ code is written in object-oriented programming manner. Portability of *NUVOLA* (between Windows-XP and many Linux operating systems such as *FEDORA 4*, *SCIENTIFIC LINUX, REDHAT9 etc.*) was accomplished in a special way of the re-switching of advanced libraries. According to the *Gaussian'03* user license, any part of the *Gaussian-03* can be included in any programming object. When the reference is made to *NUVOLA*, the Gaussian–3 task is not integrated info the *NUVOLA* content.

The main scheme of *NUVOLA* is presented as follows. Fig. 3 represents the principal scheme of *NUVOLA*, where all the thfee logical prts are presented. The original modules P81, *GAUSSIAN* and *P83* operate using advanced script system. According to the *INPUT* request, the P81 link responsible for generating starting files and scripts in order to run the *Gaussian'03* package out of the *NUVOLA* contest. *Gaussian'03* produces the matrixes of eigenvalues and eigenvectors of the separated molecules and the density matrix of the joint molecule. The *GAUSSIAN* link is used for the intermediate data reading and distributing in the memory. The *P83* link is intended for resulting the data calculation as multipurpose *OUTPUT*. The inter-modular data exchange was implemented through intermediate file system. As well, the system of control output was (realized) established for each link.

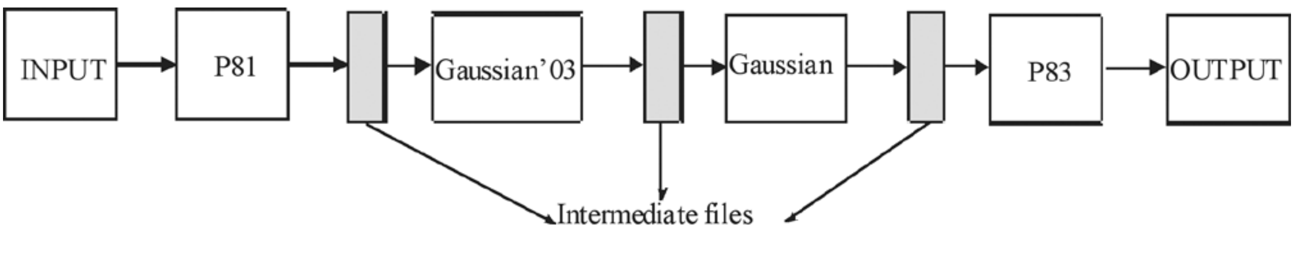


Fig 3. Principal scheme of *NUVOLA*

The *INPUT* conception was accomplished through the keyword managing system. The structure of the starting file (named conditionally as *NAME.kkk*) is presented in Fig.4. A convenient structural file indicates all features of each calculation object (MS-1 and MS-2) and operation. *NAME.kkk* is the only file which has to be made by the user. It presents a strictly determined structure, namely:
i) particular keywords handling the program;
ii) two matrixes containing structural positions of atoms in MS.

Two different styles are allowed: Z-matrix as well as Cartesian matrix. Unfortunately, combining these two styles in one **.kkk* file is not allowed. Depending on the required accuracy of further calculations in the *Gaussian'03* programs, appropriate keyword values have to be selected.

NUVOLA's KEYWORDS
Blank line
TITLE LINE
Blank line
Z-matrix or cartesian matrix of MS-1
Z-matrix or cartesian matrix of MS-2

Fig 4. Structure of *NAME.kkk* file

Fig. 5 represents the detail scheme of the P81 link. Starting from the request from the file *NAME.kkk,* preparation of several *Gaussian-03* input files (*NAME123.gjf*, *NAME234.gjf* and *NAME345.gjf* and corresponding batch file *NAME.bcf*) was done in this part, including formal verification of molecular structure and formal creation of several working scripts. The *P81* link also generates several additional input files for data handling and visual control purposes, namely:

i) *NAME.yes* - for next modules *GAUSSIAN* and *P83*;
ii) *NAME.ent* according *HYPERCHEM* standard;
iii) *NAME.mol* according *DALTON* standard.

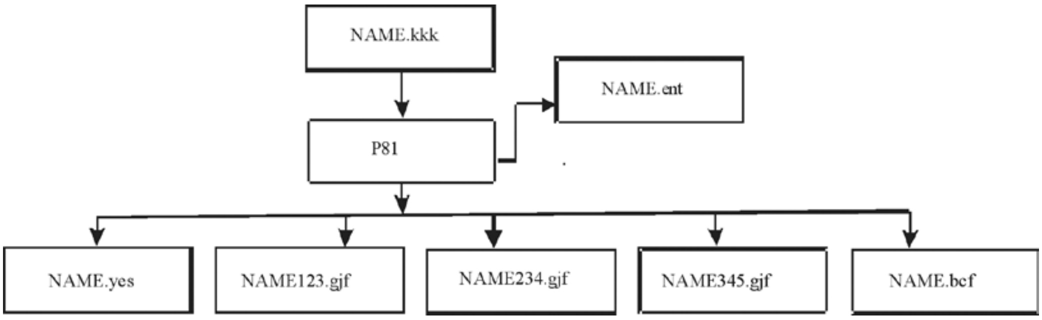


Fig. 5. Detail scheme of *NUVOLA*: *P81* link

First, let us consider the two interacting molecules. According to the general formulation of a physical task, formal decomposition of a physical object has to be accomplished. The atoms of three formal MS – MS-1, MS-2 and joint MS-12 presented in the associated coordinate system will be used (corresponds to *NAME123.gjf*, *NAME234.gjf* and *NAME345.gjf*).

Second, the ab-initio Hartree-Fock (HF) method or the electron correlation method (MP2, MP4) as well as the corresponding basis set (simple STO-3G or more appropriate 6-311G) have to be selected according to the model of a physical task. The polarization functions (from d,p up to 3df, 3pd) as well as diffusion functions for hard atoms can be included due to physical model properties. The full distribution of state population has to be output. For achieving the required result, a standard command line to run a *Gaussian03* task for all three MS should be written as follows:

#P HF/6-311G(3df,3pd) pop=full sp (7)

Fig. 6. represents the scheme of file handling in *Gaussian'03*. Depending on the request placed in command line, three single point (SP) jobs will be provided using batch script *NAME.bcf*. After the data-processing, we receive two text-type files *NAME123.out*, *NAME234.out* and one binary-packed file *NAME345.chk*. Then *NAME345.chk* has be converted using the *Gaussian'03* standard utility FormChk from the binary to text-formatted *NAME345.fch* file.

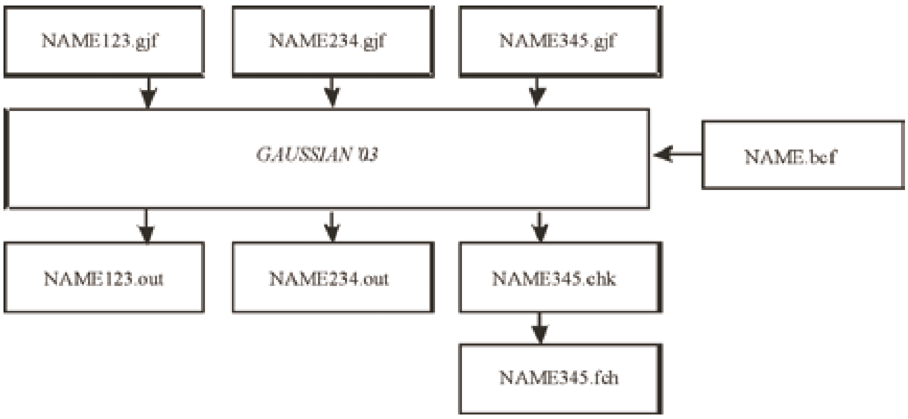


Fig. 6. File handling in *Gaussian'03*

The values of main parameters of each electronic system (eigenenergies and eigenvectors of *MS-1* and *MS-2*) were presented in the output files *NAME123.out*, *NAME234.out*. Eigenfunctions $e_{1\alpha}$, $e_{2\beta}$ are constructed as

$$e_{1\alpha} = \sum_{i=1}^{N} C_{\alpha i}\varphi_{1i} , \qquad (8)$$

$$e_{2\beta} = \sum_{k=1}^{M} C_{\beta k}\varphi_{2k} , \qquad (9)$$

where index 1, 2 denotes the number of interacting MS, and index α, β denotes the molecular orbital, which correspond to the state number. The electronic state α belongs to MS-1 and β belongs to MS-2, respectively. The values φ_1 and φ_2 express atomic orbits of the MS-1 and MS-2, and coefficients C_α , C_β corresponds to the eigenvectors of the MS-1 and MS-2, respectively (see table 1).

Table 1. Eigenvalues E as molecular orbitals (MO) and corresponding eigenvectors φ as atomic orbits (AO) of MS-α and MS-β

Two molecular systems - MS			
MS-1α		MS−2β	
Eigenvectors, AO, related to MS-α	Eigenvalue, MO of MS-α	Eigenvalue, MO of MS-β	Eigenvectors, AO, related to MS-β
N-number of particles of MS-α		M-number of particles of MS-β	
$\varphi^1_{N,\alpha}\varphi^2_{N,\alpha}\varphi^3_{N,\alpha}\cdots\varphi^N_{N,\alpha}$	$E_{N,\alpha}$	$E_{M,\beta}$	$\varphi^1_{M,\beta}\varphi^2_{M,\beta}\varphi^3_{M,\beta}\cdots\varphi^M_{M,\beta}$
$\varphi^1_{N-1,\alpha}\varphi^2_{N-1,\alpha}\varphi^3_{N-1,\alpha}\cdots\varphi^N_{N-1,\alpha}$	$E_{N-1,\alpha}$	$E_{M-1,\beta}$	$\varphi^1_{M-1,\beta}\varphi^2_{M-1,\beta}\varphi^3_{M-1,\beta}\cdots\varphi^M_{M-1,\beta}$
…	…	…	
$\varphi^1_{L,\alpha}\varphi^2_{L,\alpha}\varphi^3_{L,\alpha}\cdots\varphi^N_{L,\alpha}$	$E_{LUMO,\alpha}$	$E_{LUMO,\beta}$	$\varphi^1_{L,\beta}\varphi^2_{L,\beta}\varphi^3_{L,\beta}\cdots\varphi^M_{L,\beta}$
$\varphi^1_{H,\alpha}\varphi^2_{H,\alpha}\varphi^3_{H,\alpha}\cdots\varphi^N_{H,\alpha}$	$E_{HOMO,\alpha}$	$E_{HOMO,\beta}$	$\varphi^1_{H,\beta}\varphi^2_{H,\beta}\varphi^3_{H,\beta}\cdots\varphi^M_{H,\beta}$
…	…	…	
$\varphi^1_{3,\alpha}\varphi^2_{3,\alpha}\varphi^3_{3,\alpha}\cdots\varphi^N_{3,\alpha}$	$E_{3,\alpha}$	$E_{3,\beta}$	$\varphi^1_{3,\beta}\varphi^2_{3,\beta}\varphi^3_{3,\beta}\cdots\varphi^M_{3,\beta}$
$\varphi^1_{2,\alpha}\varphi^2_{2,\alpha}\varphi^3_{2,\alpha}\cdots\varphi^N_{2,\alpha}$	$E_{2,\alpha}$	$E_{2,\beta}$	$\varphi^1_{2,\beta}\varphi^2_{2,\beta}\varphi^3_{2,\beta}\cdots\varphi^M_{2,\beta}$
$\varphi^1_{1,\alpha}\varphi^2_{1,\alpha}\varphi^3_{1,\alpha}\cdots\varphi^N_{1,\alpha}$	$E_{1,\alpha}$	$E_{1,\beta}$	$\varphi^1_{1,\beta}\varphi^2_{1,\beta}\varphi^3_{1,\beta}\cdots\varphi^M_{1,\beta}$

The values of the main parameters of the joint electronic system (density matrix of MS-12) were presented in the output file *NAME345.fch*. Table 2 represents the value of ρ density matrix for MS-12. *N* and *M* denote the number of AO in MS-1 and MS-2, respectively.

Fig. 5 represents the detailed scheme of the *GAUSSIAN link*. *NAME123.out*, *NAME234*.out and *NAME345.fch* files (original Gaussian'03 output files) are sent to *GAUSSIAN* link as input files for further ordered formation of vectors and matrixes without any mathematical operations. These files (*.out) contain the tables of molecular orbital energy values, corresponding the atomic orbital (AO) energy, AO coefficients and density matrix coefficients for pending MS. The *GAUSSIAN* block provides:

i) unformatted input of different values;
ii) and as a result – ordered output files containing extracted energies of *MO*, *AO* coefficients – *NAME123.orb*, *NAME234.orb* – and density matrix coefficients – *NAME345.den.*

As follows, several arrays are created:

i) 1D arrays – vectors of eigenenergies for each MS-1 and MS-2;
ii) 2D arrays of eigenvectors for each MS-1 and MS-2
iii) 2D array of ρ density matrix coefficients for MS-12

Table 2. ρ density matrix for MS-12.
N and M denote the number of AO in MS-1 and MS-2, respectively.

		MS-1	MS−2
		N	M
MS-1	N	$\begin{matrix}\rho_{11} & \rho_{12} & \cdots & \rho_{1N}\\ \rho_{21} & \rho_{22} & \cdots & \rho_{2N}\\ \cdots & \cdots & \cdots & \cdots\\ \rho_{N1} & \rho_{N2} & \cdots & \rho_{NN}\end{matrix}$	$\begin{matrix}\rho_{1,N+1} & \rho_{1,N+2} & \cdots & \rho_{1,N+M}\\ \rho_{2,N+1} & \rho_{2,N+2} & \cdots & \rho_{2,N+M}\\ \cdots & \cdots & \cdots & \cdots\\ \rho_{N,N+1} & \rho_{N,N+2} & \cdots & \rho_{N,N+M}\end{matrix}$
MS−2	M	$\begin{matrix}\rho_{N+1,1} & \rho_{N+1,2} & \cdots & \rho_{N+1,N}\\ \rho_{N+2,2} & \rho_{N+2,2} & \cdots & \rho_{N+2,N}\\ \cdots & \cdots & \cdots & \cdots\\ \rho_{N+M,N} & \rho_{N+M,2} & \cdots & \rho_{N+M,N}\end{matrix}$	$\begin{matrix}\rho_{N+1,N+1} & \rho_{N+1,N+2} & \cdots & \rho_{N+1,N+M}\\ \rho_{N+2,N+1} & \rho_{N+2,N+2} & \cdots & \rho_{N+2,N+M}\\ \cdots & \cdots & \cdots & \cdots\\ \rho_{N+M,N+1} & \rho_{N+M,N+2} & \cdots & \rho_{N+M,N+M}\end{matrix}$

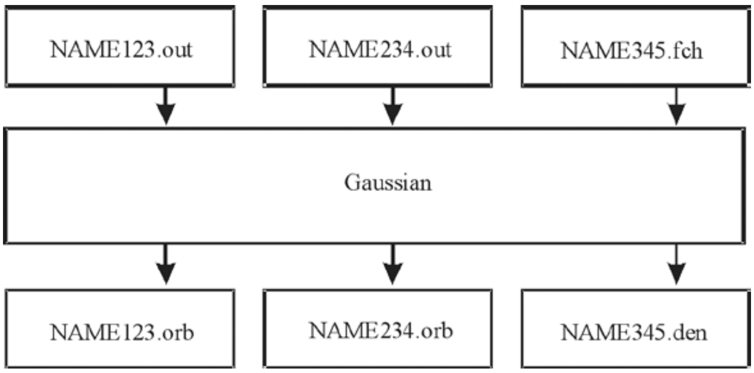


Fig. 5. The detailed scheme of *NUVOLA*: the *GAUSSIAN* link

Fig. 6 represents the detailed scheme of the *P83* link which performs the main calculation of the IET rate, according to Eq.(6). It includes the most mathematically complicated part - computing of multidimensional values of coupling S. Several files – containing the AO coefficients *NAME123*.orb, *NAME234*.orb, density matrix coefficients *NAME345*.den and initial Cartesian matrix data of the joint molecule *NAME234.yes* – are sent to the *P83* block as input files for further matrix manipulations, resulting IET rate matrixes in the *OUTPUT* files. Such set of files is labeled as W in Fig. 7.

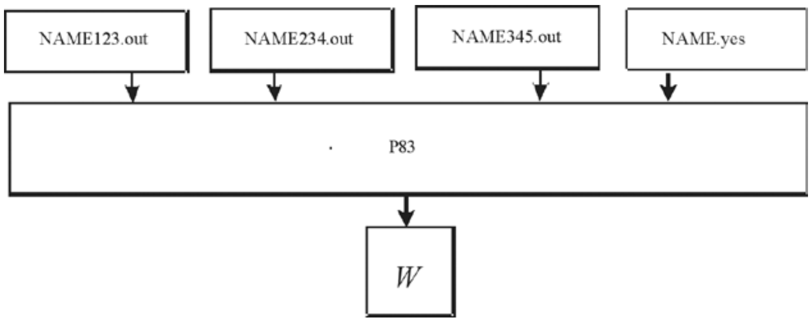


Fig. 6. The detailed scheme of *NUVOLA*: the *P83* link.

The following manipulations were performed to ρ density matrix of MS-12 presented in Table 2. The mentioned ρ density matrix contains the following four blocks:

$$\rho = \begin{vmatrix} A_1 & A_2 \\ A_3 & A_4 \end{vmatrix} \quad (10)$$

where each A_x (x=1..4) represents a different type of MS interactions. According to the general formulation of a physical task, the *intermolecular* (MS-1→MS-2) process has only to be treated and estimated, in contrast with the *intramolecular* (MS-1→MS-1, MS-2→MS-2) process (which is negligible). In that case, the formal reconstruction of ρ density matrix has to be implemented in order to exclude only intramolecular interactions. The mentioned density matrix (Eq. 10) contains four blocks which corresponds to the intramolecular interactions (diagonal blocks) as well as intermolecular interactions (non-diagonal blocks). Obviously, in that case, the diagonal matrix-blocks have to be excluded (the matrix elements of blocks A1, A4 have to be equated to zero). The newly reconstructed matrix ρ_0 contains real values in non-diagonal blocks A2, A3 only.

We have used Eq.(11) for routine evaluation of intermolecular coupling ξ_{if} :

$$\xi_{if} = \eta_i \times \rho_0 \times \mu_f \, . \quad (11)$$

where ηi , µj represent the two-block vectors of MS-1, MS-2, respectively, containing eigenvectors - coefficients related to AO. Indexes i and j represent the number of initial and final electronic states, respectively. The two-block vectors (1D arrays) are created according to the scheme presented in Eq. (12) and Eq. (13):

$$\eta_i = \underbrace{C_{i,1} \quad C_{i,2} \quad \ldots \quad C_{iN}}_{N} \; \underbrace{0 \quad 0 \quad \ldots \quad 0}_{M} \quad (12)$$

$$\tilde{\mu}_f = \underbrace{0 \quad 0 \quad \ldots \quad 0}_{N} \; \underbrace{C_{f,N+1} \quad C_{f,N+2} \quad \ldots \quad C_{f,N+M}}_{M} \quad (13)$$

4. New matrix multiplication formalism

The total transition rate W could be calculated using Eq. (2), where the intermolecular coupling $S_{\alpha\beta}$ for each transition corresponds to value ξ_{if} expressed by Eq. (11) as a product of matrix multiplication. The order of matrixes depends on the properties of the physical task. For example, $[H_2..H_2]$ derivative requires 12 members (2 MS * 2 atoms * 3 AO, using HF/6-311G basis set).

Eq. (14-15) show a construction of three matrix multiplication process, where each MS requires only two members – two-member-set of AO of the MS-1 (block A1 consist of a1, a2) and MS-2 (block C2 consist of c3, c4). First and third block-vectors contain zero-values as described above.

$$S_{ac} = \begin{vmatrix} A_1 & 0 \end{vmatrix} \times \begin{vmatrix} 0 & B_{12} \\ B_{21} & 0 \end{vmatrix} \times \begin{vmatrix} 0 \\ C_2 \end{vmatrix} \quad , \qquad (14)$$

$$S_{ac} = \begin{vmatrix} a1 & a2 & 0 & 0 \end{vmatrix} \times \begin{vmatrix} 0 & 0 & b1 & b2 \\ 0 & 0 & b3 & b4 \\ b5 & b6 & 0 & 0 \\ b7 & b8 & 0 & 0 \end{vmatrix} \times \begin{vmatrix} 0 \\ 0 \\ c3 \\ c4 \end{vmatrix} \quad , \qquad (15)$$

The two-dimensional density matrix ρ_0 (Eq. 14) contains four blocks. The two non-diagonal blocks reflect the behavior of intermolecular interactions (block B12 consist of b1, b2, b3, b4 and block B21 consist of b5, b6, b7, b8). The two diagonal blocks contain zero-values.

The classical routine of matrix multiplication is well-known [15]. The intermediate and final products of matrix multiplication were proceeded according to Eq. (16-17) and Eq. (18-19), respectively.

$$S_{ac} = \begin{vmatrix} 0 & (A_1 B_{12}) \end{vmatrix} \times \begin{vmatrix} 0 \\ C_2 \end{vmatrix} , \qquad (16)$$

$$S_{ac} = A_1 B_{12} C_2 , \qquad (17)$$

$$S_{ac} = (a1 \cdot b1 + a2 \cdot b3) \cdot c3 + (a1 \cdot b2 + a2 \cdot b4) \cdot c4 . \qquad (19)$$

The presented routine using Eq.(17-18) requires immense computer recourses- CPU time and virtual memory. This is obvious because the IET ratio *W* has to be calculated and evaluated for each transition made up from LUMO and higher states of MS–α to LUMO and higher states of MS–β. The number of all transitions N_{TOTAL} is equal to:

$$N_{TOTAL} = \left(\frac{N+M}{2} \right)^2 \quad , \qquad (19)$$

where N and M denotes the number of AO in MS-1 and MS-2, respectively. It is necessary to emphasize that the combination of very small MS requires negligible number of transitions (for example, *N=M=10*, N_{TOTAL}=100). Unfortunately, real model systems require the computing of *W* values in case, when N_{TOTAL} is equal to several thousands and more. In this case, multiplying of three matrixes according to Eq. (14) seems to be irrational, because all of them contain the blocks filled up by zero-valued elements. It means that significant excluding of zero-multiplication operations can shorten the time of *W* computation. Obviously, it is necessary to optimize the computational algorithm.

The optimization of the multiplication processes was implemented as follows:

i) to exclude the resource-wasting multiplication by zero – according to Eq.(14);
ii) to exclude the time-wasting processing of intermediate product – according to Eq.(16) ;
iii) to cancel the multiplication loop even if the value of *delta*-function is too small – according to Eq.(6).

So, it is reasonable to process only the first half of the first vector (block A_1), the quarter of density matrix (block B_{12}) and only the second half of the second vector (block C_2). Fig. 7 represents the scheme of the new matrix multiplication formalism. The assumption Y_1 requires taking into account the quarter of density matrix (block B_{12}). The assumption Y_2 requires column-row reorganization of the block-vectors A_1 and C_2

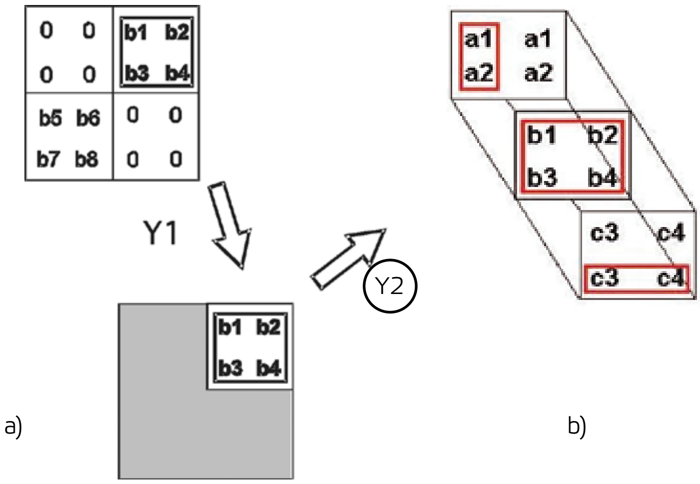


Fig. 7. Scheme of the novel matrix multiplication formalism:
a) the assumption Y_1 requires taking into account the quarter of density matrix (block B12) ;
b) the assumption Y_2 requires column-row reorganization of the block-vectors A_1 and C_2.

The formal reorganization of the classical multiplication routine described earlier was organized by means of three-dimensional matrix Ω(*N, M, 3*), where the first index N and the second index M denote the numbers of AO in MS-1 and MS-2, respectively. The third index represents the previous matrixes in the reduced form: the block-vectors A_1 and C_2 are redirected into the 1st and 3rd layer, respectively, and the block B_{12} of density matrix - into the 2nd layer. It is necessary to emphasize that the block B_{12} is presented in the previous form when the block-vectors A_1 and C_2 (used to fill the 1st layer in the column shift and the 3rd layer in the row shift, respectively (see Fig. 7b, the vector-block in presented in red). The final product was received according to Eq.(21) as direct multiplication of the matrix elements which represent the same result as in the classical case shown in Eq. (19):

$$S_{ac} = a1 \cdot b1 \cdot c3 + a2 \cdot b3 \cdot c3 + a1 \cdot b2 \cdot c4 + a2 \cdot b4 \cdot c4 \cdot \quad (20)$$

However, the number of transitions N_{TOTAL} is reduced to a significantly lower value:

$$N_{TOTAL} = \left(\frac{N}{2} \times \frac{M}{2} \right), \quad (21)$$

5. Different forms of the IET rate matrixes

Quantitative intermolecular ET processes can be analyzed in the framework of energies as well as spatial charge redistribution. Fig. 8 represents the detailed output of the P83 link.

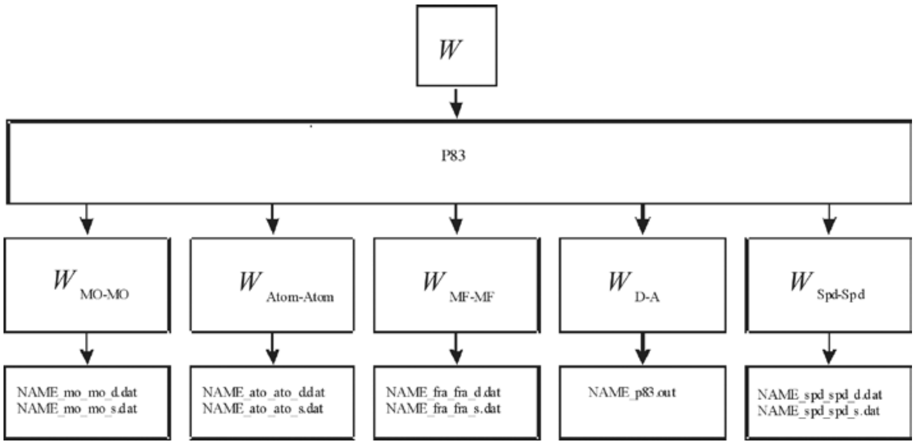


Fig. 8 The detailed scheme of *NUVOLA*: the output of *P83* link.

The estimation of the IET rate between two MS can be allowed in the frame of the molecular charge redistribution between AO. This assumption corresponds to the general task of the physical process (IET). This assumption (called as AO-AO assumption) is quite useful in case of very small MS. For example, in case of $[H_2..H_2]$ derivative, 36 transitions have to be analyzed (6*6=36, each MS contains 12 AO, but only 6 AO are of LUMO and higher). Unfortunately, it is very difficult to analyze real model systems containing hundreds and thousands atoms (it corresponds to the number of AO equal to several thousand and more).

The second assumption (called as ATO-ATO assumption) can be allowed in the frame of the molecular charge redistribution between atoms. It is quite useful in case of small MS (containing several atoms). According to the selected basis set and method, all IET rates belonging to different AO of the same atom are summed. This operation reduces the order of the IET rate matrix until the amount described by indexes (N_{MS-1}, N_{MS-2}), where NMS-1, NMS-2 are the number of atoms in corresponding MS.

The third assumption (called as FRA-FRA assumption) can be allowed in the frame of the molecular charge redistribution between molecular fragments. It is quite useful in the case of big MS (containing several hundred atoms). In the same cases, several levels of molecular organizations can be presented. For example, π-electronic system of benzene ring containing fragments as well as n-electronic system of carbonyl fragments could be formally treated as molecular fragments, which play significant role in the charge redistribution processes. In that case, both previously formally decomposed MS according to a theoretical request could be analyzed in the third assumption. According to the list of formal decomposition into molecular fragments, all the IET rates belonging to different AO of the same molecular fragment are summed. This operation reduces the amount of the IET rate matrix until amount described by indexes (NFRA-1, NFRA-2) , where NFRA-1, NFRA-2 are the number of molecular fragments in corresponding MS.

The fourth assumption (called as D-A assumption) can be allowed in the frame of the molecular charge redistribution between MS-1 (donor) and MS-2 (acceptor). It is useful if we want to estimate IET in a molecular stack (partially) and the behavior of the IET process (overall). This operation represents the IET rate as a single number.

The fifth assumption (called as SPD-SPD assumption) can be allowed in the frame of the molecular charge redistribution between the same types of AO. It is useful when analyzing the mixed ET (for example, between *s* and *p* orbitals, between *s* and *d* orbitals etc). According to the selected basis set and method, all the IET rates belonging to an AO of the same type are summed. This operation reduces the order of the IET rate matrix until the dimension 3*3 (in case of *spd* type) or 4*4 (*spdf*) etc.

All the required matrixes are taken to make the resulting data files and three-dimensional molecular electrons redistribution diagrams.

Conclusion

The quantum chemistry program package *NUVOLA* was prepared and tested. The estimation of IET between two molecular systems can be allowed in the frame of the charge redistribution between:

i) atomic orbitals;
ii) molecular orbitals;
iii) atoms;
iv) molecular fragments;
v) donor and acceptor pairs.

Quantitative intermolecular ET processes could be analyzed in the framework of energies as well as spatial charge redistribution.

Acknowledgements

This work is part of *KARSKOPAS* research program which is financially supported from Lithuanian State Fundation for science an education. Nurija Galikova acknowledges the Lithuanian Science Council for the Summer Undergraduate Research Fellowship Award (2007).

References

[1] Frank Jensen. Introduction to computational chemistry - Wiley, Second edition, 2007.
[2] Bolton J.R., Archer M.D. Basic Electron - Transfer Theory // In: Bolton J.R., Mataga N., McLendon G. Electron Transfer in Inorganic, Organic and Biological Systems.- *Advances in Chemistry Series*, CSC Symposium Series 2. - American Chemical Society, Washington, DC 1991. - Pp. 7-23.
[3] Rodrigo E. Palacios, Gerdenis Kodis, Christian Herrero, Ernesto Mariño Ochoa, Miguel Gervaldo, Stephanie L. Gould, John T. M. Kennis, Devens Gust, Thomas A. Moore, Ana L. Moore. Tetrapyrrole Singlet Excited State Quenching by Carotenoids in an Artificial Photosynthetic Antena *// J. Phys. Chem. B*, 110 (50), 2006. - 25411 -25420
[4] Resonance energy transfer. Ed. David L. Andrews, Andrey A. Demidov. - Chichester: John Wiley and sons Ltd., 1999. – P. 320.
[5] Volkhard May, Oliver Kühn. Charge and energy transfer dynamics in molecular systems, 2000 – Pp. 249-333.
[6] Scholes, G.D., and Ghiggino, K.P.. Electronic interactions and interchromophore energy transfer. // In: Advances in Multiphoton Processes and spectroscopy . S. H, Lin, A. A. Villaeys, and Y. Fujimura, editors. World Scientific, Singapore, Vol. 10, 1996. Pp. 95-331.
[7] Arieh Warshel, William W. Parson. Dynamics of biochemical and biophysical reactions: insight from computer simulations *// Quarterly reviews of Biophysics*, vol.34, 4 (2001) P. 563-679.
[8] A.V. Barzykin, P.A. Frantsuzov, K. Seki, M.Tachiya. Solvent effects in nonadiabatic electron-transfer reactions: theoretical aspects. *// Advances in Chemical Physics*, Vol. 123. 2002. – P. 511-616.
[9] N. Galikova, A. Gruodis, M Balevičius. Quantum chemistry program NUVOLA. – Vilnius university, 2007. 3rd ed.
[10] R. A. Marcus. On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. V. Comparison and Properties of Electrochemical and Chemical Rate Constants *// J. Phys. Chem. 67*, 853, 2889 (1963)
[11] R.A. Marcus and N.Sutin *// Biochim.Biophys.Acta 811* (1992) 265
[12] Nurija Galikova, Alytis Gruodis, Mindaugas Leonas Balevičius, Leonas Valkunas Intermolecular electron transfer due to quantum chemistry approach. I. General theory and calculation routine // in preparation.
[13] Nurija Galikova, Alytis Gruodis, Mindaugas Leonas Balevičius, Leonas Valkunas Intermolecular electron transfer due to quantum chemistry approach. II. Processes in stack of α-crystal of N, N-dimethilaminobenzylidene indan-1,3-dione // in preparation.
[14] Gaussian 03, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 2003.
[15] Knuth, D.E. Seminumerical Algorithms // In: The Art of Computer Programming. Vol. 2. Third Edition, 1998. - pp 501.

VILNIUS BUSINESS COLLEGE

Vilnius Business College
Kalvarijų str. 125, Vilnius, LT-08221, Lithuania
Tel./ fax. +370 5 215 4884, Mob.: +370 616 55422, +370 615 96680