Big Data: An Exploration Toward the Improve of the Academic Performance in Higher Education

Leonardo Emiro Contreras Bravo(✉), Giovanny Mauricio Tarazona Bermudez(✉), and José Ignacio Rodríguez Molano(✉)

Universidad Distrital Francisco José de Caldas, Bogotá, Colombia
{lecontrerasb,gtarazona,jirodriguez}@udistrital.edu.co

Abstract. An outlook is presented, summarizing the types of information analysis that are applying currently to the educative area like the: Data mining, academic analysis and learning analytics, identifying the advantages of the implementation, the use of the methodology b-learning and big data in the pedagogical processes. Later, an approach of the existing problem in the Colombian university is made, with the purpose of propose an application model tools of big data to determinate student retention, desertion and university academic performance.

Keywords: Analysis of data · Big data · Virtual environments
Engineering education

1 Introduction

In the last years, the use of diverse informatics tools applied to the education area, has been perceived. Some technologies has allowed the integration of the traditional teaching and the virtual, such as the case of the LMS (Learning Management System) used in different Colombian universities with the purpose of search us the better teaching methods, [1] and they have become in a student academic information important repository, in which the last one interact, in such a way that it is suppose that the process teaching – learning, mediated by the technologies is favoring the student, becoming it in an active actor of the process. One of the most important factors in the efficiency of the educative process is the retention parameter, defined as the difference between the number of students that enter in the first semester and the graduated students per year [2]. Likewise, the academic performance is the success main indicator or student failure and the factors that affect the abandonment possibility of a higher education program is multivariate [3]. Therefore, its determination has generated controversy, because there is no definitive theory about a methodology for its measurement or an indicator for its valuation. Being multidimensional, the academic performance depends of various aspects such as the teacher, institution, student objectives, between others [4].

Even more, in matter of the academic performance in the higher education, most of the important research present a marked interest in the inclusion of personal factors in the analysis of the existing empirical evidence, there are few the studies that make an approach multivariate that include variables from the approach "learning analytics or

Y. Tan et al. (Eds.): DMBD 2018, LNCS 10943, pp. 627–637, 2018.
https://doi.org/10.1007/978-3-319-93803-5_59

academics" [5]. Some recent works has applied statistical methods with an approach generally based on demographic, economic, psychological factors, such as are the researches of [6, 7], between others.

Regarding the topic, in the environment of the education in Colombia has been evidenced in the last years, a significative increase in the levels of desertion by the students in courses of higher [8]. So much so, that the study of this problematic has been started from different universities like the Antioquia and the Andes, who have taken as reference the academic information content in the Information System SPADIES (open information) that allows establish different categories by which it can present the student desertion. However, the virtual teaching platforms such as WEbCT, Moodle, Blackboard, Claroline, Dokeos and recently the platforms MOOC (Massive Open Online Courses) have allowed to have in the universities that use it, the option of watch over in real time the student's activity. The integration of this information with others variables, is constituted in the origin of the concept of "learning analytics [9]. Therefore, being the academic analytics and the analytics of the learning emerging fields, in the case of the Colombian education, the culture of use and analyze the information generate by the student in the formation processes and learning to determinate its influence in the academic performance, the desertion or the graduation is not developed.

2 Generalities

2.1 Big Data

It can be defined like a concept that make reference to the recollection storage and treatment of a set information too big with the purpose of find repetitive patterns inside them, that result impossible control them with the tools database and conventional analytics. It does it using different techniques, between them the data mining. Through the definition of models and the use of the different technologies it's wanted become the information in a active of great value. The technologies sector of information and communication has been leader in the trend to control a huge quantity of information, that are generating in real time and they come from different medias such as the social networks, sensors, technology platforms and a lot of audio and video devices that require use that information in the statistical report creation and predictive models that can be used in many areas of the human job.

It could be named some of the possible applications of the Big Data in different areas. One of this, is the medicine, where more clearly is being drawing this benefit, because the treatments and the personalized medicines are, without a doubt, a desirable future to everyone. Likewise in the security area and defense of its use is strengthen more the security against the organizations attack and terrorists, others [10].

2.2 B-Learning

Through the use of the platforms LMS, it has been introduced to the educative area the term B-learning, defined like that learning mode that mix the teaching by traditional activities with non-face-to-face technology: "which combines face-to-face and virtual

teaching" [11], taking advantage of the best of the virtual teaching and the face-to-face one, through the use of electronic resorts or digitals available on internet [12]. In the education area, the teachers also desire to evaluate their courses and have doubts like, for example, if that students that practice with their exercises online continuously are better in the final exam that the students that does not? With this and other questions can be useful the learning tools, collecting, analyzing and visualizing automatically the correct information [13]. However, it could be thought that this could through the tools of the virtual learning environments (VLE); which, this is not completely true, because most of watch over and generation tools, these last are designed to collect, analyze and visualize information in a tabular static way that was pre-designed by the system developers. Therefore, the teachers face to the difficulty of using tools of Learning Analytics that help to answer their individual questions continuously and efficient facing the missing tools in VLE.

2.3 Data Mining

This technique that also is used as information raw material, tries to discover patterns, trends, profiles or others relations that can be of interest, that are present in the information volume to analyze, but keep hide. Some of the methods that are used to the information analysis are: artificial intelligence, automatic learning, statistics and information base systems. This are used when the information is stored in conventional systems relational type (Data Warehouse); in the opposite case when the information is from heterogeneous information or not structured that come from different source, then is necessary use technologies of the Big Data. It can be affirmed that there is a certain relation between this terminology. Data Mining require of Big Data to streamline its processing and information management, as well as Big Data require of Data Mining to the predictive analysis of information and to can detect trends

3 Learning Panorama Improved by Technologies

The educative area has been characterized by the incursion a Little late of the technologic application, in the Big Data's case, it doesn't look like be the exception [14]. Therefore, researchers and teachers have proposed seep in the interest of the technology use to maximize the learning experience, this is, the potentized learning by the technology with the technology. The Big Data incursion in education has originated the rise of the TEL (Technology enhanced learning), and a series of concepts, tightly related, which have as fundamental characteristic, the fact of capture a series of educative information that are analyzed to get useful information that allows take previous actions of the teaching or activities to foment teaching and learning. The new concepts emerging (EDM - Educational Data Mining), Academic Analytics, Learning Analytics (Fig. 1). The learning analysis is defined according to the society for learning analytics research (SoLAR), as the measurement, recollection, analysis and information divulgation about the students and their contexts, with the purpose and understand and streamline the learning and the environments in which is produced.

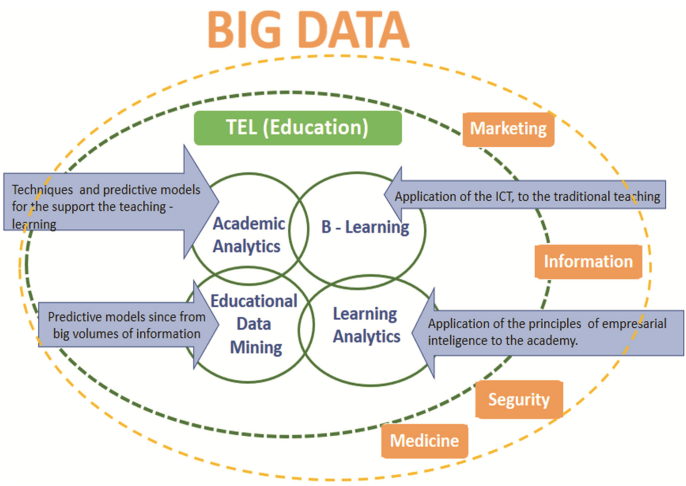

Fig. 1. TEL panorama – learning improved by technologies.

The EDM approach is focused more specifically in the development, the research and the application of methods computerized to detect languages patterns in big educative information sets; that is to say, the information reduction and analysis of those individual components and the relations between them. Meanwhile, the academic analytics, strengthens the responsibility and the objective fulfillment of the higher education institutions, allowing them work efficiently with the resorts, whose purpose is improve the aspects such as: the academic performance, decrease the desertion rate and increase the graduation rate. According to [15], statistical analysis and prediction models with the purpose of this mix institutional information create awareness in the students, teachers o administrators to can change the academic behavior. So much so, that They use the term "academic analytic" as the equivalent of "business intelligence", to describe the intersection between the technology, the information, the management culture and the information application to manage the academic company. In the other hand, the learning analytics use analytics techniques and predictive models to support the institutional and curricular objectives, offering to the student support resources like support to the achievement of learning objectives. Meanwhile the EDM focus specifically in make an analysis of a reduced set of information for in that way, allow supporting the decision making concerning to the learning and teaching model in the educative institutions. In conclusion, the academic analytics approach are the institutions and educative systems, while the analytics learning focuses its attention in help the students and teachers to raise the learning objectives. So that, the academic analytics and the learning analytics are focused specifically in tools and methods to the information exploration coming from educative contexts, which makes to regard current how the techniques that will help to mold the higher education future and generate new approaches and strategies in learning and teaching learners improvement [13].

4 Taking Decision Impulse by the Big Data Analytics

So that, the Big Data application in the higher education can help to develop new, learning methods, making easy or personalizing the learning and making easy the generation of new knowledges, as well as making easy the early disorders detection [16], learning difficulties and desertion problems and academic performance. It should be mentioned that in this area in Latin America and specially in the Universidad Detrital in Colombia, where this research work starts. There are economic resources limitations for the implementation of this technology, as well as human resort (teacher) trained, added to the infrastructure [17]. There are planned diverse stages that have the process actors (teachers, students, and institution) who through the application of different information analysis techniques, could help to improve the learning-teaching process and thus the academic performance, decreasing the desertion.

4.1 Learning Analytics Approach

Is a Big Data action area, that has dabbled in the education, making easy through the internet and the learning management systems (LMS), having like information and collection sources, the information took in the online learning systems, such as: platform Moodle access, discussion forums, conclusion, activity evaluations and its interactions with the purpose of analyze them and propose improve actions of the learning process. It should be mentioned that according to the NMC Horizon Report: 2016 Higher Education Edition, which purpose is identify and describe the technologies that will have a great impact in the education around the world in the next five years; the learning analytics is a trend to be adopted in the higher education in a year, such as companies do it like Amazon, Netflix y Google in the marketing management and affiliate users, in where that can help to transform the education, going from a system of standard of only size to a flexible system, designed to satisfy the needs and academic interests of students. Starting from the analytic process planned by [18], authors like [19] generate the life cycle of Learning Analytics (Fig. 2), which consist in four main stages.

The first one named learning environment, involves various systems LMS, such as Moodle, which will be used to extract some information. The second one named Big Data, consist in educative information of the different groups (interaction with LMS, personal information, traces or clicks made in the LMS, between others) that can be processed. The third one named analysis, understands different analytical methods as the statistical analysis that search discover hide patterns interesting in the educative information set. Within quantitative analytical methods, are the statistical analysis, that search predict with trust what is going to happen in the future to can take decisions more intelligent applying stats and math operations. Another option, is the social networks quantitative analysis which allows identify relations between them (teachers and students). The fourth one is get the objectives and streamline the learning environment (prediction, recommendation, personalization, reflection, feedback).

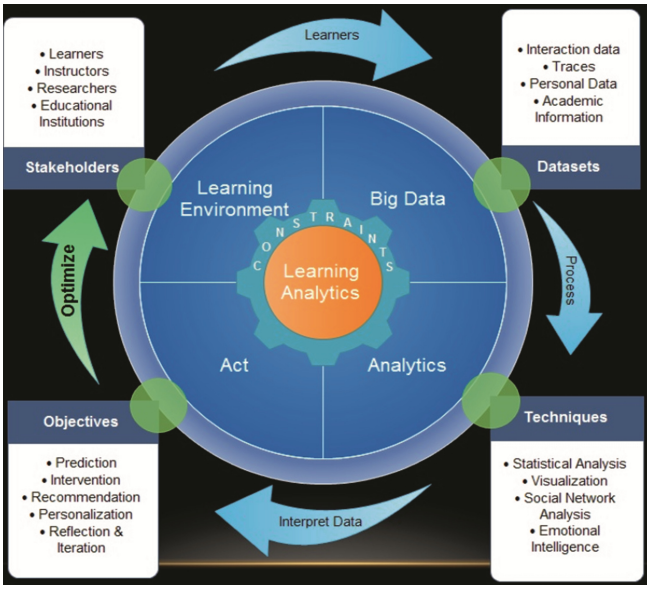

Fig. 2. Life cycle of the Learning Analytics Source: [19]

4.2 Academics Analytics Approach

The academic analytics focus specifically in tools and methods to the information exploration that come from educative contexts, with the purpose of generate new approaches and strategies in the teaching improvement and the learning of the learners [13]. Different authors, among them, agree in that, the teaching analytics offers like potentialities that could help to increase the school retention, predict the student progress, streamline the learning and teaching techniques, also it improves the decision making in the educative institution [20]. Thus, the same way that the learning analytics, the academic analytics makes use of the information that the data mining gets through the pattern searching or application of data processing algorithms to strengthen the responsibility and fulfillment of the higher institutions objectives, allow them work efficiently with the resources, whose purpose is get the aspects such as the improvement of the academic performance, decrease the desertion rate and increase the graduation rate [21]. However, the technology to deliver these potential is very young yet, and the research in the understanding of the pedagogical usefulness is in its childhood yet [22]. In the Table 1, there are shows some of the potentialities of these techniques according to the actor to regard in the educative ambit:

Table 1. Interest groups in AA, LA and its benefits. Adapted from: [19, 23]

Interest group	Objectives	Examples
Students	Improve their performances Personalize the online learning Personalize the learning activities according to the student profile Recommend courses	The students are informed about the learning process and compare their performance with others, They start early the task and make questions giving signals through the applications
Teachers/Researchers	Search behavior patterns Improve its teaching methods Sort out or group students in base to their characteristics Explore a number value/ continuous degree to the performance Make interventions on time Change the aptitude to recommendations to the students based in their activities Evaluate the past work to improve the future experience y become it in learning according to the personalization and adaptation	Control the learning progress of the students using applications Through the visualizations, the researchers of the course can compare the analysis techniques learning to can recommendation actions Prevent the school desertion, determinate what students may be at risk, advise the students that may need additional assistance and improve the success Recommend a discussion, suggest a course or recommendation books related to what previous students consulted Evaluate courses, improve the course models, discover new deliver methods of educative information
Educative institutions	Organize the institutional resources Improve the Schedule design Improve the efficiency of the portals Support the processes of decision to get higher educative goals	Increase the retention rate, control the higher education objectives, through the retention rate increase, using signals that provide the applications.

5 Project with Big Data

The integration of a great software quantity or platforms that store a lot of information that come from the student is proposed, use the open information ready by government entities and universities, between others with the purpose of try it, analyze it and visualize it with the encouragement of support the decision making. So, it contributes to the

information reutilization giving place to applications much richer that contribute more knowledge or that complete the existing information. The challenge is in process huge information quantities and become them in intelligent and timely decisions [24].

It is planned to start the project various steps, which are show in the Fig. 3: Initially the aspects that pretend resolve with Big Data technologies. In second place, is necessary collect educative information through the different platforms available to the university, open information, data base, and public sources. In the same way a Learning Analytics process start, in which is required to acquire information, through the student interaction (data producer) with different activities stablish in the Moodle platform (LMS). In third place, is necessary process previously the information to get significate information, the initial information must to be infiltrate and cleaned for later be stored.

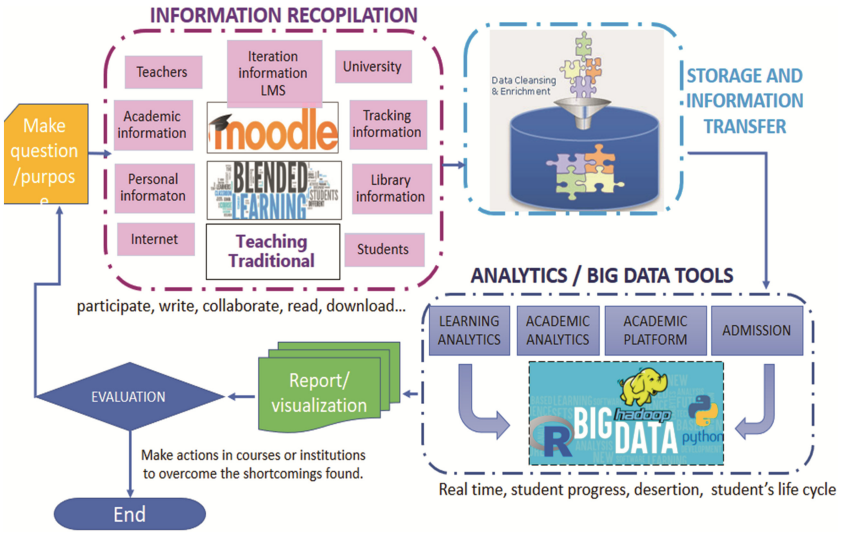

Fig. 3. Project scheme with Big Data, Adapted of [25]

Later an analysis of the same by the Big Data tools is made. A gamma of statistical methods are applied to the same information including the prediction model (used to interfere information such as the abandonment rate and learning results) social networks analysis (that analyze the relation between the people networks, groups and organizations), the mining relation (that search identify patterns between the information set, such as the student success rates) and the information for the human judgment (information visualization that allows to the teachers give an timely answer to the students and also to take decisions based in the evaluation of educative information [20].

It could be thought that this analysis could be made with the traditional way for the information analysis that are based mainly in mathematical models, and the information is collected to adjust to the models. So, it could be planed the paradigm`s solution to the problem impulse by the information. The approach based in information is not only centered in predict what is going to happen, but also focus in what is happening in this moment and how to be propagate to the future events. The science information

(statistical and computing fusion) to use, will fulfill 2 objectives [26]. Prediction: to predict the answer/exit of variables of future entry; and inference: to deduce the association between the answer variables and the entry variables. Inside of the application possibilities of algorithms to solve a problem, it is not eliminated the use of evaluative computing algorithms (EC), which one is one of the computing and artificial intelligence branches that is applied to the resolution of combinatorial optimization problem. Those are a technique searching and optimization set [27], between the most known are probably the genetic algorithms that look for the best possible solution to a problem trying to imitate what is happening in the environment. So, the algorithms could be used to resolve problems of analysis Big data in the real world. As is planned [28], the future of science information, is in mix the strengths of the algorithms EC and the Big data analysis to design news algorithms in the optimization or information analysis.

Finally, the analytical results are presented to the final user, as reports, online recommendations or through the decision making. The information visualization, through the graphics and tables can make that the analytical results be easy to interpret and understand. If the results doesn't have sense, it must to reconfigure the problem and restart the process [25]. This job faces an important challenge and is the analysis of the different techniques and technologies used to the information analysis, required open architecture solutions able to capture information (structured and not structured), important information and administrative and academic systems from the context. If the different information can be integrated, it would get an enriching square of the student life cycle from the initial admission, going through the student learning experience until arrive to the graduation and employment [18].

6 Conclusions

This work approaches new concepts that can be regarded as a component that helps to improve the education through the use of technology and communication. Concepts that has be applied with success in other areas and that when is applied to the educative area integrating the different information bases, platforms or systems that surround the student in only one, can easy the learning process, maybe through, a personalize education that can generated starting to the analysis of characteristics or patterns presented by each student in particular.

In the university student life cycle a considerable quantity of information that has little bit use is generated. Those can be integrated through an application designed to be used by different types of teachers. Both for the beginner and the expert that require to resolve specific questions about the learning of a student, its group or perhaps take actions from the university institution point of view, that has a specific question and wants to make a deeper analysis. Likewise, not all the information is useful to the integration process of Learning Analytics and the others technologies, therefore, it will have to define, which is the important information of all that can be collected? What aggregation level is needed? The analysis information frequency and the actions to undertake, will be on real time or conversely will be make each hour, each day, each month or each year?

It pretends design a tool or software with the main purpose of support the reflection and the improve of the learning methods based in interest and personal observations. Because is required the learning analysis tools desired to provide a clear, simple, easily to interpretable information, with a flexible interface for the information, information exploration, as well as visualization of evaluations results based in graphic indicators selected individually.

References

1. Schroeder, U.: Web-based learning – yes we can! In: Spaniol, M., Li, Q., Klamma, R., Lau, R.W.H. (eds.) ICWL 2009. LNCS, vol. 5686, pp. 25–33. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03426-8_3
2. Salcedo, A.: Desertion in Colombian Universities. Rev. Acad. y virtualidad **3**(1), 50–60 (2010)
3. Estrada, J., Quintero, R., Vasquez, D.: Bajo rendimiento académico en la Universidad Distrital Francisco José de Caldas, Primera. Editorial UD, Bogotá (2015)
4. Khan, I., Choi, J.: An application of educational data mining (EDM) technique for scholarship prediction. Int. J. Softw. Eng. Appl. **8**(12), 31–42 (2014)
5. García, D.: Construcción de un modelo para determinar el rendimiento académico de los estudiantes basado en learning analytics (análisis del aprendizaje), mediante el uso de técnicas multivariantes, February 2016
6. Tomás, J., Expósito, M., Sempere, S.: Determinantes del rendimiento académico en los estudiantes de grado. Un estudio en administración y dirección de empresas. Rev. Investig. Educ. **32**(2), 379–392 (2014)
7. Kotzé, M.: Psychological resources as predictors of academic performance of first-year students in higher education., Sielkundige bronne as voorspellers van Akad. prestasie van eerstejaarstudente die hoëronderwys **45**(2), 85–121 (2013)
8. Guzmán, C.: Deserción estudiantil en la eduacion superior colombiana, Ministerio de eduacion nacional- Republica de Colombia, p. 158 (2009)
9. Ejreaw, A., Drus, M.: The challenges of massive open online courses (mooc) – a preliminary review. ICOCI Kuala Lumpur. Univ. Utara Malaysia, pp. 25–27 (2017)
10. Carrillo, J., et al.: Big Data en los entornos de Defensa y Seguridad, Inst. Español Estud. Estratégicos, vol. 1, p. 124 (2013)
11. Dipro, E.J.C., Almenara, J., Díaz, M.: ICT training of university teachers in a Personal Learning. **1**(1), 2 (2012)
12. Picciano, A.: Big data and learning analytics in blended learning environments: benefits and concerns. Int. J. Interact. Multimed. Artif. Intell. **2**(7), 35 (2014)
13. Dyckhoff, A., Zielke, D.: Design and implementation of a learning analytics toolkit for teachers. Educ. Technol. Soc. **15**(3), 58–76 (2012)
14. B. I. G. Data and E. N. La, "Big data," vol. 17, pp. 1–16 (2016)
15. Baepler, P., Murdoch, C.J.: Academic analytics and data mining in higher education. Int. J. Scholarsh. Teach. Learn. **4**(2), 17 (2010)
16. Argonza, J.: Big data en la educación. Rev. Digit. Univ. **17**(1), 1–16 (2016)
17. Nunes, J.: Estado da Arte sobre Analítica da Aprendizagem na América Latina. no. Cbie, p. 1024 (2015)
18. Elias, T.: Learning Analytics: Definitions, Processes and Potential (2011)
19. Khalil, M., Ebner, M.: Learning analytics: principles and constraints. EdMedia World Conf. Educ. Media Technol. **2015**(1), 1789–1799 (2015)

20. Campbell, D., Oblinger, J.: Learning analytics in Higher Education Academy. https://www.heacademy.ac.uk/enhancement/starter-tools/learning-analytics#snapshot-logo
21. Iten, L., Arnold, K., Pistilli, M.: Mining real-time data to improve student success in a gateway course. In: Eleventh Annual TLT Conference (2016)
22. Daniel, B.: Big Data and analytics in higher education: Opportunities and challenges. Br. J. Educ. Technol. **46–5**(0007–1013), 904–920 (2015)
23. Fournier, H., Sitlia, H.: The value of learning analytics to networked learning on a personal learning environment. In: LAK 2011 Proceedings of the 1st International Conference on Learning Analytics and Knowledge, pp. 104–109 (2011)
24. Duan, L., Da Xu, L.: Business intelligence for enterprise systems: a survey. IEEE Trans. Ind. Informat. **8**(3), 679–687 (2012)
25. Huang, T., Lan, L., Fang, X.: Promises and challenges of big data computing in health sciences. Big Data Res. **2**(1), 2–11 (2015)
26. Donoho, D.: 50 years of data science. Princeton (2015)
27. Kennedy, J.: Swarm Intelligence, Handbook of Nature-Inspired and Innovative Computing, pp. 187–219. Kluwer Academic Publishers, Boston (2006)
28. Cheng, S., Liu, B., Shi, Y., Jin, Y., Li, B.: Evolutionary computation and big data: key challenges and future directions. In: Tan, Y., Shi, Y. (eds.) DMBD 2016. LNCS, vol. 9714. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40973-3_1