METHOD TO CLASSIFY PAPERS TO BEGINNING STUDENTS USING NATURAL LANGUAGE PROCESSING: AN EMBEDDED PROCESSOR CASE STUDY

Tiago dos Santos Patrocínio
Federal University of Piauí/Brazil
tiagodsp93@gmail.com

Ivan Saraiva Silva
Federal University of Piauí/Brazil
ivan@ufpi.edu.br

Abstract: **In the academy, there is an untold amount of papers, and yet more new ones are presented in different fields of science. It is common in academia influence young students, beginners in the field of research, choose the source articles better known (by reputed universities or researchers) or more cited. However, recently published papers did not have enough time to consolidate itself in the scientific community, thereby having few citations, therefore hindering the search for an interesting paper. To prevent such influence and work around the problem mentioned, in this paper, we discuss a method for classification of articles using their abstracts applied in algorithms to Natural Language Processing, seeking to provide to the beginning student, an interesting read addressing the topic searched.**

Keywords: **Abstract; Readability; Embedded processor; quality; Paper; Tag; Knowledge; Beginner; Students; Natural language processing; Papers; Interesting; Quality.**

I. INTRODUCTION

The field of publication of papers is utmost important for the consolidation of scientific community. By this way all the knowledge of particular area are approached and discussed. In academy, beginning students in the field of research, are indicated and strongly influenced to search for articles by reputed universities or researchers. It is an obvious and common practice. Therefore, are disregarded numerous papers, which can be good in the sense of content, but does not highlight out for lack of reputation of its components. Furthermore, the growing number of scientific papers is imminent, resulting in an untold amount of them. These facts hinder, for beginner students, an evaluation of the recent papers published by various new authors, which disqualify potential important subjects.

In this paper we discuss a method to classify papers, analyzing the content of the abstract using Natural Language Processing, in order to avoiding the influence of reputation, providing interesting and quality papers. To do it, we analyzed recent published papers, in this case only published in 2010 to 2015 period. We choose the field of Embedded Processor, because this area fits in aforementioned situation, and at the same time is our specialty, serving as a useful example in our experiment.

In section II is defined the quality sought in our research, explaining elements used in our method. Section III shows some stats and charts about the collected data. Section IV explains our method and goals aimed. Section V is the conclusion and futures expectations.

II. QUALITY DEFINITION

Along will be mentioned the term “quality”, which is our purpose with this paper. This quality refers to the quantity of potential information stored in each paper worked in our research, which are highlighted by sentences containing in its abstract.

In the field of paper publication, the abstract is very important. It must have contain the amount of enough information which reader needs to select the paper, containing clear words and comprehensive sentences about the context subject

A. Tags:

The term quality in our discuss is measured by a set of words called “tags” which are extracted of each paper in its abstract and, then, seek for appearances of them in the following method which will be presented in section.

We used the online tool VocabGrabber [1], provided by Thinkmap Visual Thesaurus. The tool extracts a set of words from the text to calculate the relevance comparing how frequently the text uses this words versus how they are used in written English overall. We adopt as tags the common words used in scientific context, also provided by the tool.

B. Flesch-Kincaid Readability Test

To complement our method, we employ a readability test, which defines how much clear is the text, measuring the comprehension difficulty of the same [2]. We suppose that a paper abstract which approach many themes, being at the same time simple and comprehensive language, become more interesting. Therefore, the Flesch-Kincaid Readability Test fits in our goal.

To our analysis, we use the Flesch Reading Ease Score (FRES). The test utilizes a mathematical process to assess the input text by means of the number of words, syllables and sentence numbers. To measure score was used the Edit Central online tool [3]. The explanation of how test works is not the aim of our paper, and then we will be superficial in this question. The test utilizes the mathematical equation in (1).

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right) = FRES \quad (1)$$

In the original purpose of this test, the scores illustrated in the Table I were adopted as reference around the world.

TABLE I. SCORES REFERENCES

FRES Score	Notes
90.0 – 100.0	easily understood by an average 11-year-old student
60.0 – 70.0	easily understood by 13- to 15-year-old students
0.0 – 30.0	best understood by university graduates

III. COLLECTING DATA

In our project, we selected one of the most reputed Digital Libraries to collect some stack of data, containing information of article itself. The data collected are title, abstract, publisher, authors and publication's vehicle.

In order to extract paper's data, a program based on C# language was used to perform parsing and collecting metadata. The program searched the keyword "Embedded Processor" obtaining 67 results from the ACM Digital Library [4] in the year 2013 and 2747 results from IEEE Xplorer [5] in the 2010 to 2015 period.

The search done in the ACM Digital Library [4] used the keyword Embedded Processor filtered by papers published only in 2013, obtaining more than a thousand results. The search engine considered keywords Embedded and Processor separately, adding many results, explaining the high amount. Therefore, we considered to use "Embedded Processor" quoted, so the number of results falls to 67 results, explained by the direct relation of article and the subject Embedded Processor. The program developed by our team treats the URL of search result HTML page, processing each item found, requesting metadata and information of inner nodes and HTML tags to retrieve all data

The Engineering Village [6] search tool provided the paper's information in IEEE Xplorer, only extracting BibTex package from papers.

For purposes of simplicity, we use data from ACM Digital Library to explain the method in this section and in section IV. Discussion of the Section V presents IEEE Xplorer data.

The figures 1 and 2 show some stats of worksheet containing the information of the 68 articles captured from ACM. Figure 1 presents a set of paper's conferences, which are linked with the "Embedded Processor" field. In figure 2, it can be observed that some conferences have highlighted due to the significant number of articles related to the field, because they are more focused in this theme.

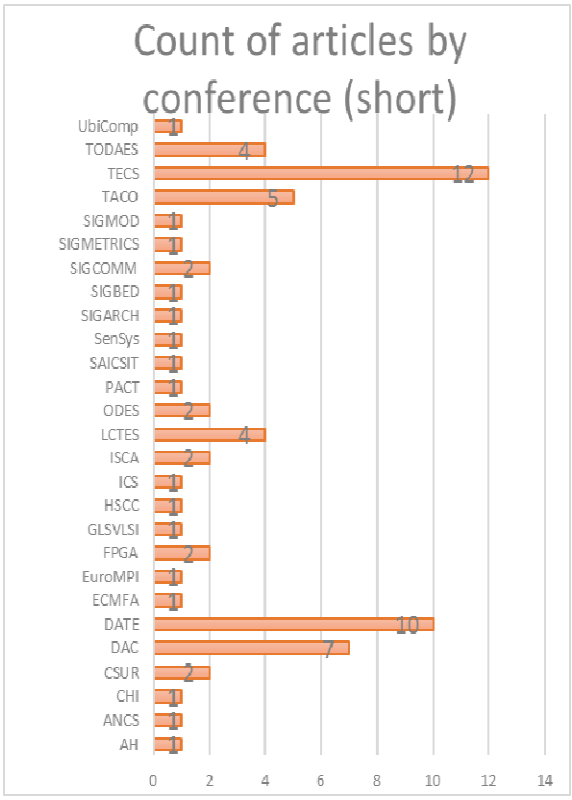

Fig. 1. Count of papers captured classified by conference (short).

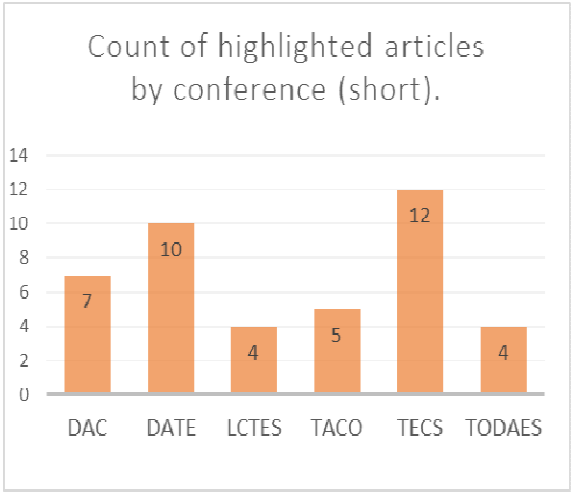

Fig. 2. Highlighted papers captured classified by conference (short).

A. *Tag extraction*

For analysis of the subject of the paper and knowledge trend, we consider to use tags in their classification. In this paper, Knowledge trend is where all paper's subjects leans converging. To extract this, we used the extraction of tags over the abstract of each paper. The process is descripted above in section II. Figure 3 shows the quantity of tags encountered in all papers illustrated in Word Cloud.

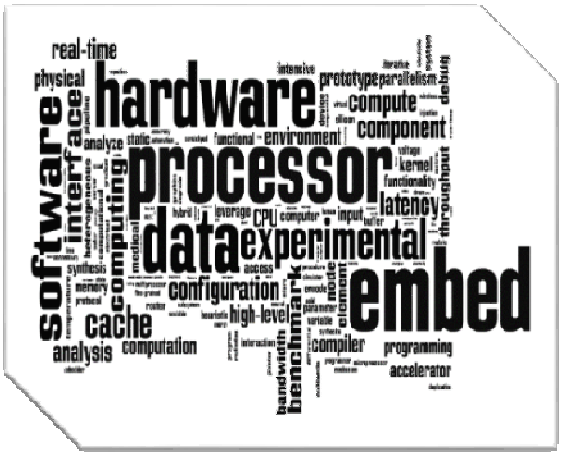

Fig. 3. Illustration of quantity of tag encountered in all papers in the worksheet.

Words like "processor", "hardware", "data" and "embed" was strongly mentioned in the subject of the papers, considering that they are covered by Embedded Processor field. However, the words "experimental", "real-time", "benchmark", "latency", among others, are highlight significantly, giving us an idea of the convergence of subjects among the papers.

B. *Pondering Tags*

To allow classification is necessary to ponder each tag. To do it, in our research we count every tag present in all papers, summed and divided by the total of tags.

Each tag uses the concept of percentage to ponder itself. Therefore, we take the number of determined tag present in all papers and then divided by the total of tags, which we call "Tag Points". Table II shows an example with data merely illustrative.

TABLE II. TAGPOINTS EXEMPLE:

TAG	SUM	Tag Points
data	22	22/62 = 0,35
cache	11	11/62 = 0,17
embed	29	29/62 = 0,46
	TOTAL: 62	

Appling to each paper in our data worksheet, their respective Tag Points, summing the value as shown in Table III. In this example, Paper02 is highlight. Figure 4 shows the result Tag Points obtained in our research.

TABLE III. SUM OF TAGPOINTS EXAMPLE:

Paper	TAGs	Sum of Tag Points
Paper01	embed	0,46
Paper02	data cache	0,52

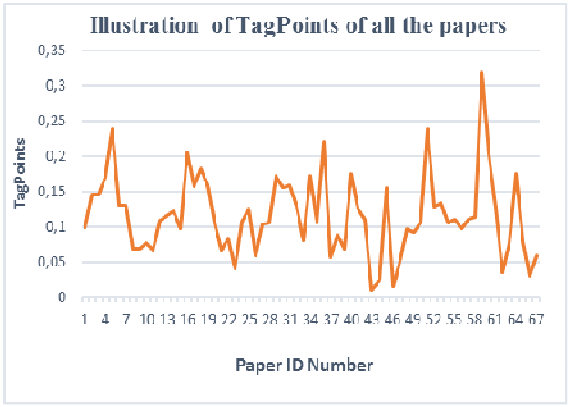

Fig. 4. Illustration of TagPoints by Paper.

C. *Readability Level:*

In order to improve our method, was decided to apply the Flesch-Kincaid test which the purpose is assess the degree of text readability. There are two types of the same test. For preference, we used the Flesch Readability Ease Score (FRES). For this test, how much more points more readable. (For more information, see section II of this paper.)

To evaluate this, was applied the test on the abstracts of the papers, because there contains the first information that the reader needs for choose an article to read. Figure 5 shows the score FRES by paper.

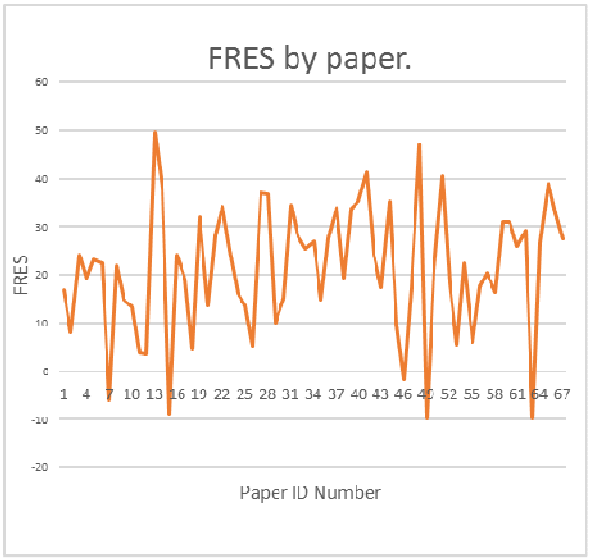

Fig. 5. Flesch Readability Ease Score by Paper.

IV. QUALITY CLASSIFICATION METHOD:

In this paper, we propose a method to classification of papers, raising to top those that obtained the best score. To do that, our classification adopts as criterion

a relation between Tag Points and Flesch-Kincaid Readability Ease test, discussed earlier.

Our intention is classify papers determining a high degree of interest, assuming that determined paper have some interesting terms in its abstract (by means of Tags), which covers the searched field, in our case, "Embedded Processor" field.

The use of tags fits in this case, because, we search by each abstract what are the most frequent terms covered by everybody in scientific field, then those that approach a high number tags, becoming more interesting among them, getting highlighting. Using the formula in (2) we can evaluate according to the above cited.

$$\frac{\mathbf{TagPoints} + \frac{\boldsymbol{FRES}}{\mathbf{100}}}{\mathbf{2}} = \boldsymbol{Score}$$

Observing the formula 2, we can see that it is an arithmetic average of the two scores, highlighting articles like as proposed above. Figures 6 and 7 shows the results in our data.

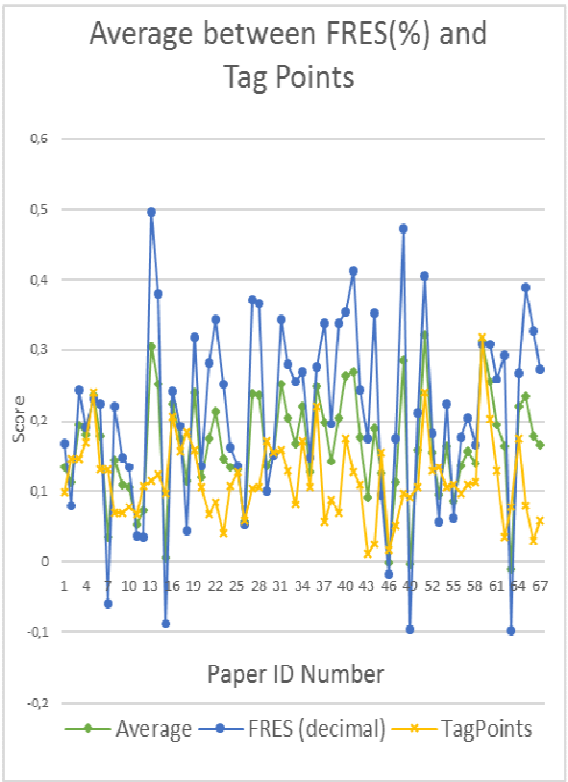

Fig. 6. Relation between TagPoints score and FRES (Flesch Readability Ease Score) sorted by paper ID number.

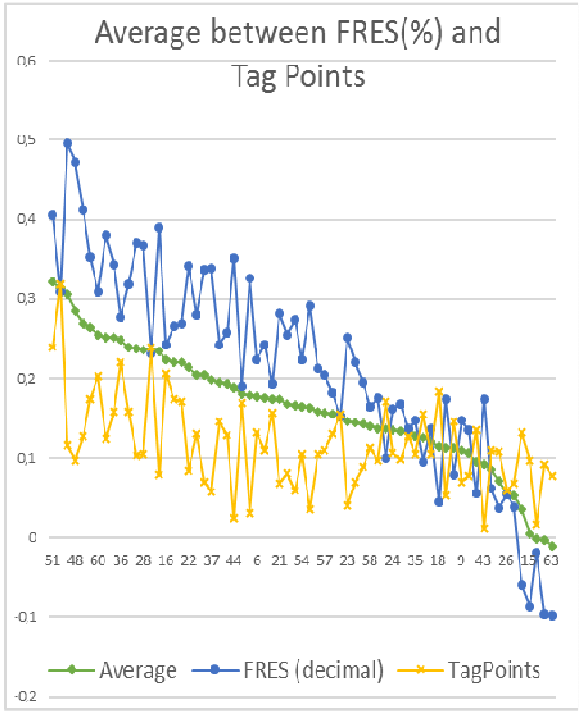

Fig. 7. Average between TagPoints score and FRES (Flesch Readability Ease Score) score in decrement order of the average.

The focus of this method is classify recently published papers, with goal of pondering one with each other based in a commons thinking or ideas provided by its abstract. The Readability test is used for rearrange the result, starting from the supposition that less complex texts with more information have much for offer for reader, becoming more indicated.

A. Citation Count

Obviously, that how much more one paper is cited in determined scientific community signifies that this one is very important, therefore, more indicated. However, it is not applied in our methodology, because the aim is the recently published papers, which not count with time to be recognized in the field, or have not reputation elements mentioned in section I.

B. Relating our Method with Citation count

To do the follow analysis, we requested papers published in the year 2010 from ACM Digital Library, because has enough time for them be recognized by the community (in this case, Embedded Processor field). Therefore, various papers has been cited by the field, providing sufficient data to us.

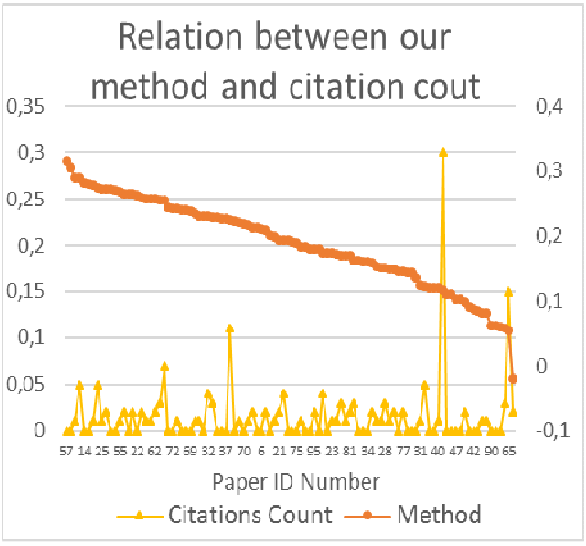

Fig. 8. Relation between our method and citation count ordered by our score method. (Data referring to the year 2010.)

As can be observed in figure 8, does not have a direct relation with our method and citation count. Papers with low classification in our method can have high number of citations and vice versa. Both manners of classification does not cancel each other. It is up to the user choose what is the best way.

C. *Results of our Method*

Was extracted 10% [7 – 13] of the total papers captured sorted by our method in Embedded Processor research field of the ACM in 2013.

V. METHODS RESULTS AND DISCUSSION

Analysis over extracted data can offer some interesting information about the method. The data is divided in 6 years in order to provide a better view of analysis. The 2015 has less data due to the unfinished year which data were extracted. Table IV shows the amount of paper data of each year.

TABLE IV. COUNT OF PAPERS BY YEAR FROM DATA USED IN RESULTS

Year	Count of Papers
2010	500
2011	500
2012	500
2013	500
2014	500
2015	247

A. *Visualization Count*

Figures 9, 10, 11, 12, 13 and 14, shows the three best papers [14 - 31] according to methods classification in their respective publication year and their amount of visualization over time. Each of best three papers presents a huge amount of visualization in their respective publication year and then occurs a decrease in the number in subsequent years. Trend lines plotted in the graphs shows a huge decrease of over the years. Therefore concluding that they are forgotten or less looked for in matter of time.

It is natural that a paper become less visualized over time, so it reinforces that the method can aid those subjects to increase their number of views by suggesting them to reader.

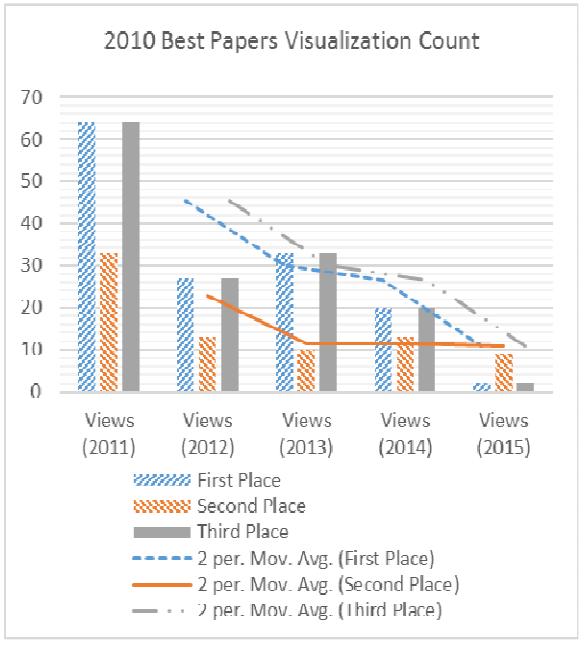

Fig. 9. Visualizations amount over 2011 to 2015 of the Best three 2010 papers according with method's classification.

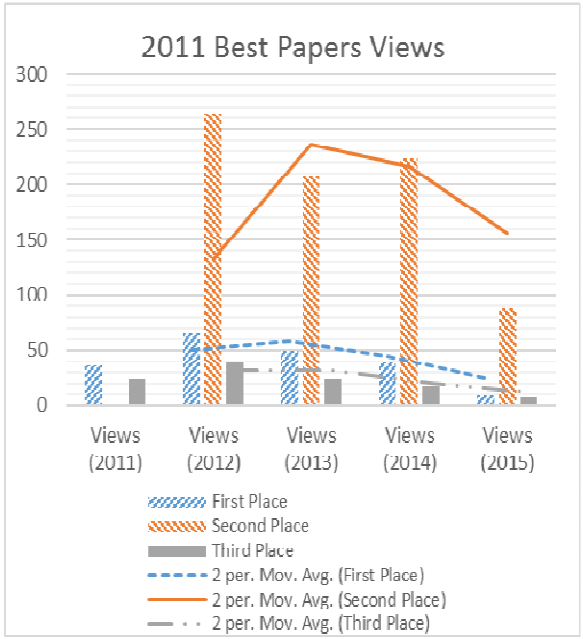

Fig. 10. Visualizations amount over 2011 to 2015 of the Best three 2011 papers according with method's classification.

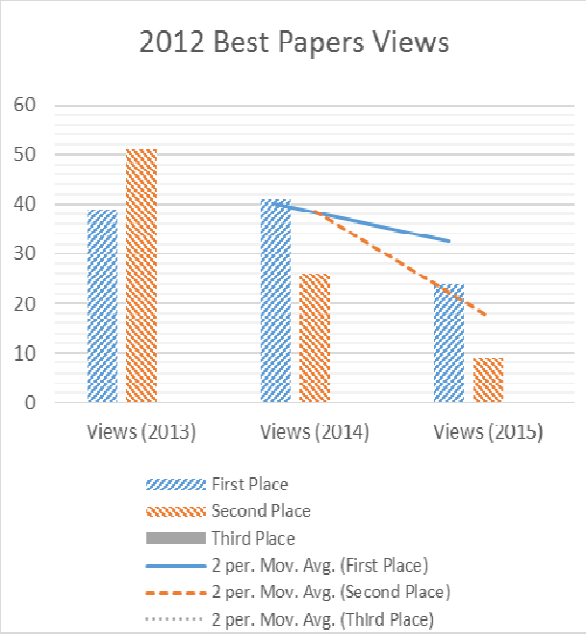

Fig. 11. Visualizations amount over 2013 to 2015 of the Best three 2012 papers according with method's classification.

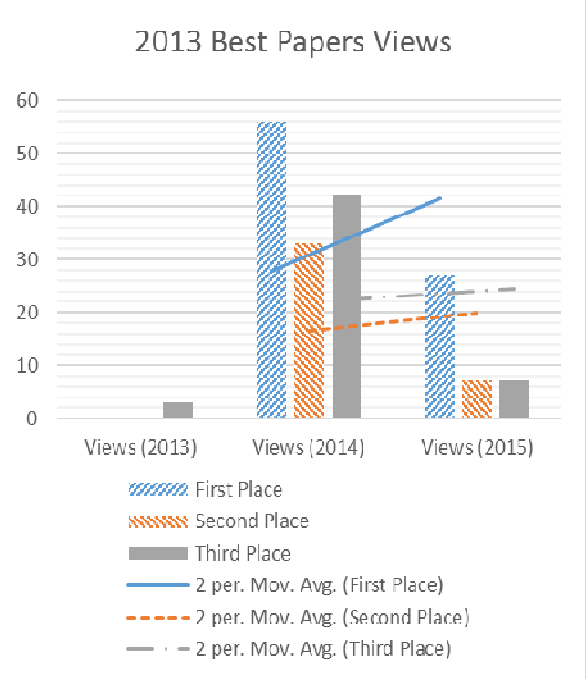

Fig. 12. Visualizations amount over 2013 to 2015 of the Best three 2013 papers according with method's classification.

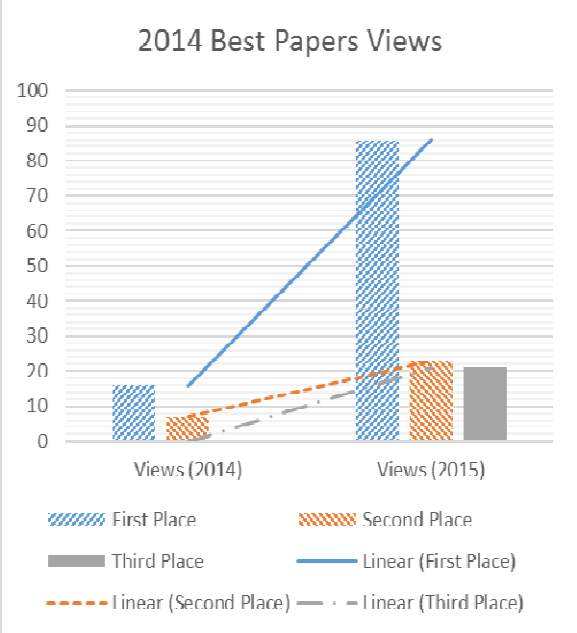

Fig. 13. Visualizations amount over 2014 to 2015 of the Best three 2014 papers according with method's classification.

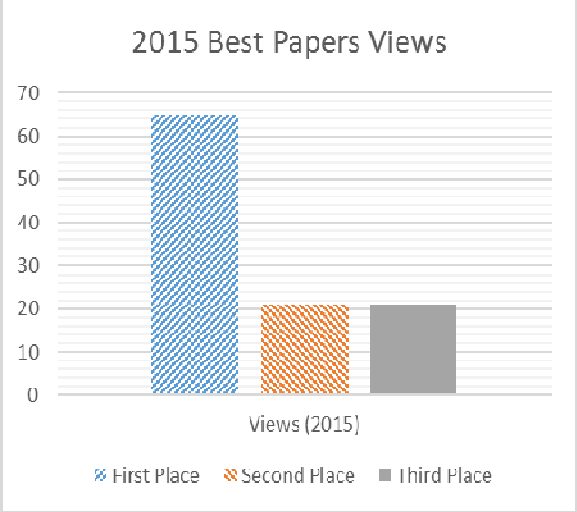

Fig. 14. Visualizations amount over 2015 of the Best three of same year papers according with method's classification.

B. Avareg by Year

Figure 15 shows average of method's classification score by year. Although the year 2014 presents the most interesting papers according with the method, in relation with the other years have a little variation in classification. The year 2015 have a little score average due being an unfinished year when data was extracted. This concludes that papers can still offer good contents about the searched context.

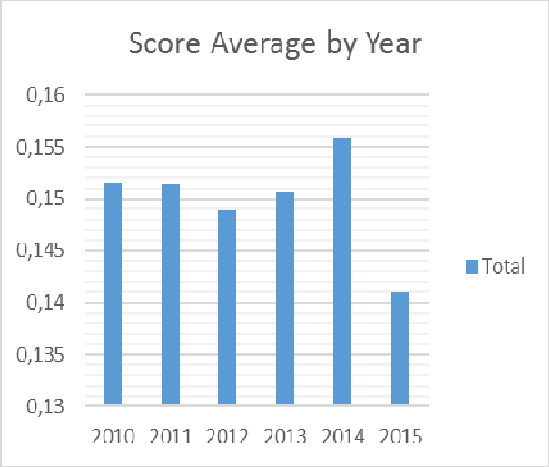

Fig. 15. Score average by year.

C. *Conferences Highlight*

Figure 16 shows the best 10 conferences in method's classification in a period of 5 years (2010-2014). The top ten classifies according to average score of method's classification achieved by papers in each conference. This means that the method can also indicate interesting conferences, which the best classified approaches the subject in a better and simple way to beginning students.

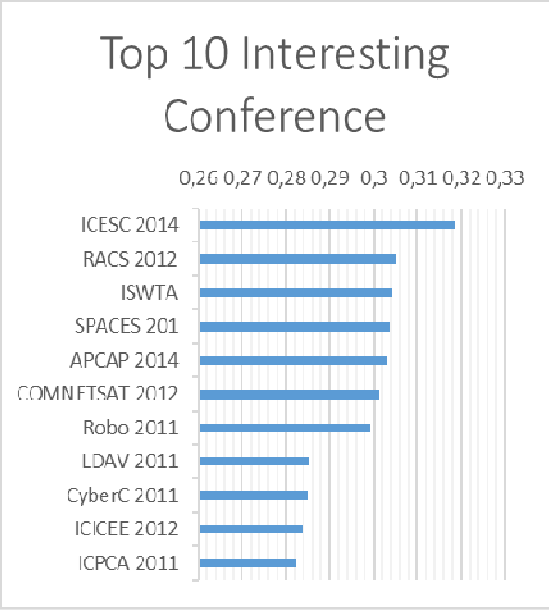

Fig. 16. Top 10 interesting conferences (2010-2014) based in average paper score.

D. *Citation Count*

The graph in figure 17 shows the average citation count and plots the average of the method's classification grouped by year. The records of papers published in early years present more citation than recent years due the longer time to be recognized in the academic community. Even with lower number of citations, the method can classify recent papers with high scores indicating the independence of the method with the paper reputation. Therefore showing that the purpose of avoid reputation is successfully achieved.

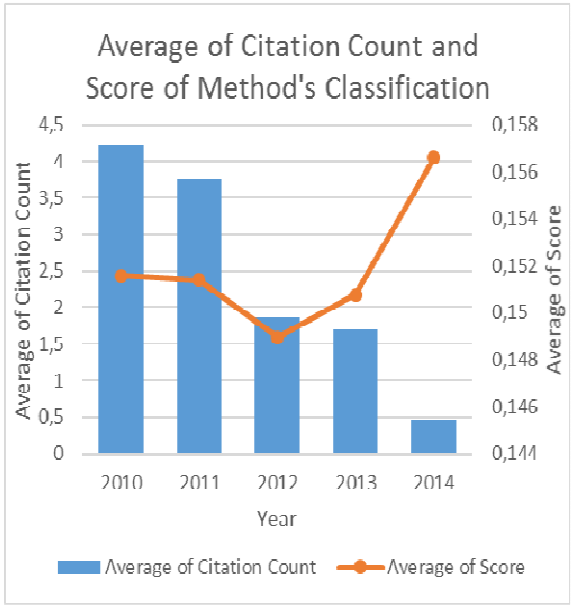

Fig. 17. Average of citation count and method's classification by year.

Figure 18 and 19 presents an overview of all paper's database in a period of 6 years (2010 to 2015) used in this study, comparing citation count and method's classification. High cited papers appears in peaks in the graph, but are not well classified by the method. This reinforces the method's reputation independence. The three more cited papers [32 - 34] has more than 70 citations.

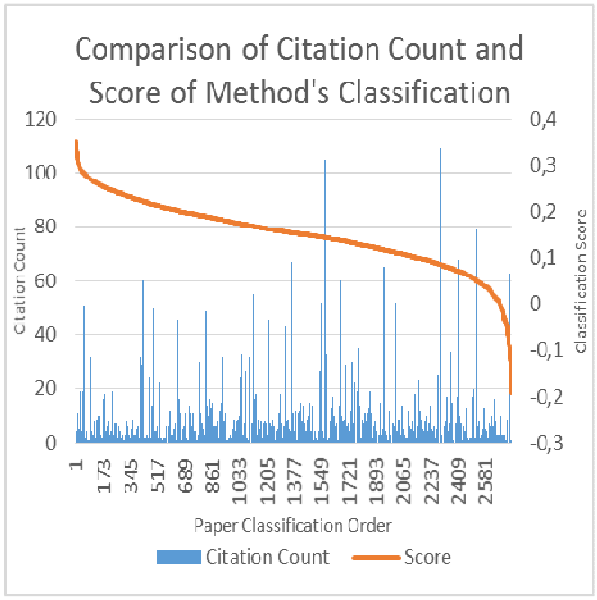

Fig. 18. Comparison between citation count and classification score of the method.

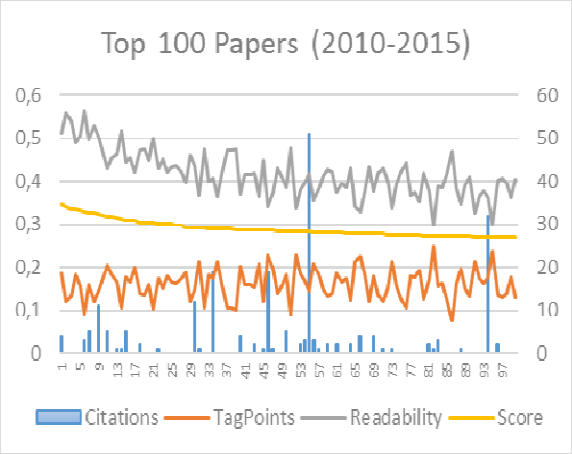

Fig. 19. Relation beteewen citation count and mathod's classifciation from the top 100 papers from 2010 to 2015 period.

E. Tags Overview

Fig. 20. Word Cloud of the extracted tags from database used to classify the papers.

Figure 20 presents a Word Cloud extracted during classification, giving an idea of the main subjects in the Embedded Processor study field over the years 2010 to 2015. The words size is proportional with frequency that they occur one time in each paper. Figure 21 presents a graph with occurrence count of some words in Embedded Processor field.

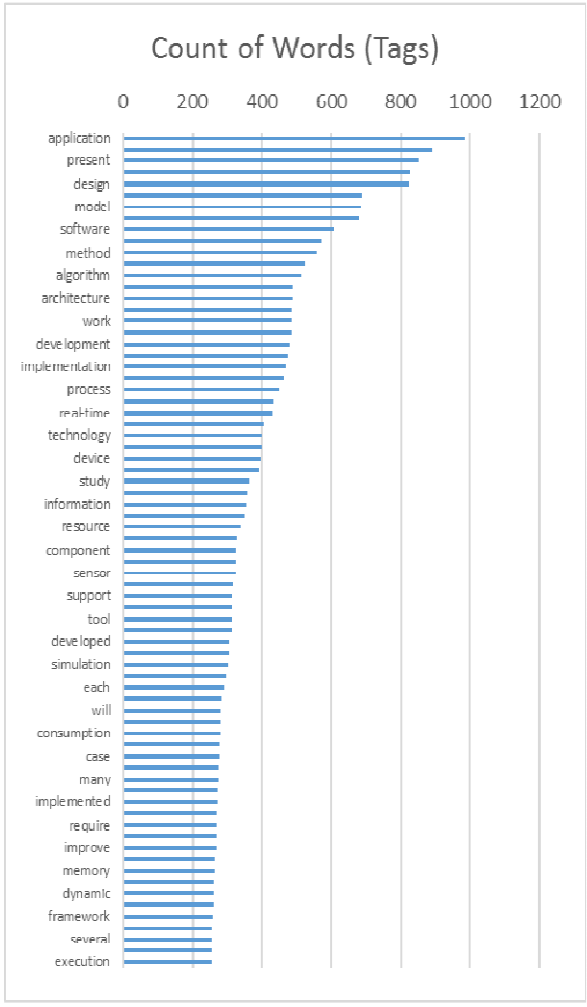

Fig. 21. Count of tag occurrences in paper's database used by the method.

VI. SURVEY

To evaluate the efficiency of our method, we performed a survey in various Computer Science communities in order to measure the interest level of the researchers regarding the papers sorted by the method.

The worst and the best papers sorted in the method's classification were mixed, so the volunteer assesses each abstract with Likert scale based on its read interest as degree without knowing the classification of each.

The survey is divided in three sections. The first section and second section contained the abstracts of the papers mixed, where only the second section had extra information regarding the author and institution of paper. The third section contained only one question, which asks the interviewee the relevance degree of this extra information in the selection of papers as interesting.

By contain only the abstract without extra information regarding the article, the first part had a behavior in favor of the method, where we had 100% success in classifying interesting and not interesting articles. The extra information in the second part has

caused a notable divergence among respondents, making 50% of them assess contrary the expectations.

In the first section of the survey, even with disagreement among some of the respondents, we noticed that most of them showed interest in relation to articles provided in accordance with the classification method. Proving that the method has great predictive ability of researchers to interesting content.

When asked in the third section of the survey, half of respondents judged relevant the information of authors and institutions of the articles necessary for the choice of interest in reading. Figure 9 shows that 70% of them had doctor degree as formation and 30% postgraduate, characterizing researchers that had formed opinions and favoritism regarding authors or institutions long their careers.

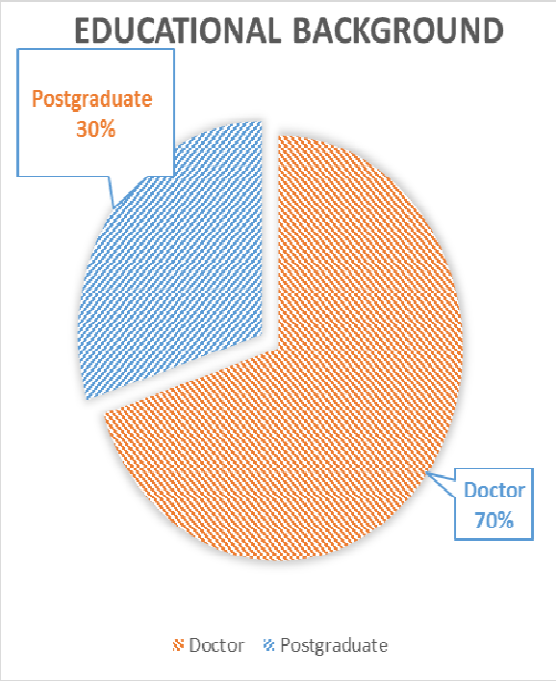

Fig. 22. Educational Background of the interviewees.

Therefore, it is remarkable that those abstracts with information alienate and cause greater disagreement over the decision on the results of the second part of the survey. However, without the information we conclude that the method is able for effectively predict interesting papers from only their abstracts in the searching scientific context, thus avoiding alienated ratings, seeking to provide only quality articles from the content provided.

VII. Conclusion and Future Expectations

Aiming provide papers for new students in research field, we propose in this paper a method that can classify them by interesting themes, only analyzing each abstract with Natural Language Processing techniques. Therefore, avoiding common influence of selection of papers by amount of citation, researcher or university reputation.

Sorting from most to least indicated in our classification, the 10% first well highlighted, has covered high relevance topics in Embedded Processor field over the years, such as: "task and data placement", "reducing intercommunication latency and power overhead", "data management for software managed multicores", "energy minimization" and "synthesis of networks". Proving that our method is able to select important and interesting themes.

In future work, we aim to utilize tools [35] developed by our expert research center in Natural Language Processing to analyze the connectivity of sentences of the context of the abstract, in order to upgrade our method. To improve the experience, we expect a better interaction between student and search engine, allowing the user to ponder the results based in reputation and various other factors.

We hope this method can assist to expand horizons, qualifying important subjects of various research in the various scientific communities, thus aid to improve and highlight new ideas in scientific field.

References

[1] VocabGrabber. Thinkmap Visual Thesaurus – 2013. Available in: < https://www.visualthesaurus.com/vocabgrabber/> Access in: 26 Nov. 2013.

[2] STEPHENS, Cheryl. All about readability – 2000. Available in: <http://plainlanguage.com/newreadability.html> - Access in: 25 Nov. 2013.

[3] Edit central – 2012. Available in: <http://www.editcentral.com/gwt1/EditCentral.html > Access in: 25 Nov. 2013.

[4] ACM Digital Library. ACM – 2013. Available in: http://dl.acm.org/ Access in: 26 Nov. 2013.

[5] IEEE Xplore Digital Library – 2015. Available in: http://ieeexplore.ieee.org/. Access in 08 Jul. 2015.

[6] Engineering Village – 2015. Available in: http://www.engineeringvillage.com/. Access in 08 Jul. 2015.

[7] Karl Viring, Sangheon Lee, Yeongon Cho, Soojung Ryu, and Bernhard Egger. 2013. Application task and data placement in embedded many-core NUMA architectures. In Proceedings of the 10th Workshop on Optimizations for DSP and Embedded Systems (ODES '13). ACM, New York, NY, USA, 37-44. DOI=10.1145/2443608.2443618 http://doi.acm.org/10.1145/2443608.2443618

[8] Pavlos M. Mattheakis and Ioannis Papaefstathiou. 2013. Significantly reducing MPI intercommunication latency and power overhead in both embedded and HPC systems. ACM Trans. Archit. Code Optim. 9, 4, Article 51 (January 2013), 25 pages. DOI=10.1145/2400682.2400710 http://doi.acm.org/10.1145/2400682.2400710

[9] Jing Lu, Ke Bai, and Aviral Shrivastava. 2013. SSDM: smart stack data management for software managed multicores (SMMs). In Proceedings of the 50th Annual Design Automation Conference (DAC '13). ACM, New York, NY, USA, , Article 149 , 8 pages. DOI=10.1145/2463209.2488918 http://doi.acm.org/10.1145/2463209.2488918

[10] André M. DeHon. 2013. Location, location, location: the role of spatial locality in asymptotic energy minimization. In Proceedings of the ACM/SIGDA international symposium on Field programmable gate arrays (FPGA '13). ACM, New York, NY, USA, 137-146. DOI=10.1145/2435264.2435291 http://doi.acm.org/10.1145/2435264.2435291

[11] Jaeyoung Do, Yang-Suk Kee, Jignesh M. Patel, Chanik Park, Kwanghyun Park, and David J. DeWitt. 2013. Query processing on smart SSDs: opportunities and challenges. In Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (SIGMOD '13). ACM, New York, NY, USA, 1221-1230. DOI=10.1145/2463676.2465295 http://doi.acm.org/10.1145/2463676.2465295

[12] Chen Huang, Bailey Miller, Frank Vahid, and Tony Givargis. 2013. Synthesis of networks of custom processing elements for real-time physical system emulation. ACM Trans. Des. Autom. Electron. Syst. 18, 2, Article 21 (April 2013), 21 pages. DOI=http://dx.doi.org/10.1145/2442087.2442092 http://doi.acm.org/http://dx.doi.org/10.1145/2442087.2442092

[13] Chen Huang, Frank Vahid, and Tony Givargis. 2013. Automatic synthesis of physical system differential equation models to a custom network of general processing elements on FPGAs. ACM Trans. Embed. Comput. Syst. 13, 2, Article 23 (September 2013), 27 pages. DOI=10.1145/2514641.2514650 http://doi.acm.org/10.1145/2514641.2514650

[14] H. N. Mishra and Y. K. Patel. Design, simulation and characterization of memory cell array for low power sram using 90nm cmos technology. IEEE Joint Societies Chapter of IE/PEL/CS, Allahabad, India, 2010.

[15] X. Zhang, Y. Tie, and D. Li. Design and realization of an embedded storage system based on lpc2387 microprocessor. volume 8, pages V8663 – V8666, Shanxi, Taiyuan, China, 2010.

[16] S.-S. Lu, C.-H. Lu, and P.-A. Hsiung. Congestion- and energy-aware run-time mapping for tile-based network-on-chip architecture. volume 2010, pages 300 – 305, Taichung, Taiwan, 2010.

[17] S. Lee, B. Lee, K. Koh, and H. Bahn. A demand-based ftl scheme using dualistic approach on data blocks and translation blocks. volume 1, pages 167 – 176, Toyama, Japan, 2011.

[18] S. Ruj, A. Nayak, and I. Stojmenovic. Dacc: Distributed access control in clouds. pages 91 – 98, Changsha, China, 2011.

[19] T. Hanawa, T. Boku, S. Miura, M. Sato, and K. Arimoto. Pearl and peach: A novel pci express direct link and its implementation. pages 871 – 879, Anchorage, AK, United states, 2011.

[20] V. Bhatia, M. Goel, S. Gupta, P. Iswerya, N. Pandey, and A. Bhattacharyya. Low power delay proficient current mode adc design. pages ABB – Power and Productivity for a Better World –, Allahabad, India, 2012.

[21] V. Bhatia, M. Goel, S. Gupta, P. Iswerya, N. Pandey, and A. Bhattacharyya. Low power delay proficient current mode adc design. pages ABB – Power and Productivity for a Better World –, Allahabad, India, 2012N. Yoo, S. Yang, and H. Jeong. A u-tour system for a tour leader in a guided group package tour. pages 291 – 296, Bandung, Indonesia, 2012.

[22] Y. K. Lee, H. Park, and C. Jeon. Fast booting based on nand flash memory. pages 451 – 452, San Antonio, TX, United states, 2012.

[23] W.-C. Chou, W.-Y. Lin, M.-Y. Lee, and K. F. Lei. Design and assessment of a real-time aceelerometer-based lying-to-sit sensing system for bed fall prevention. pages 1471 – 1475, Manchester, United kingdom, 2013.

[24] A. Siblini, E. Baaklini, H. Sbeity, A. Fadlallah, and S. Niar. Efficient FPGA implementation of h.264 cavlc entropy decoder. Marrakesh, Morocco, 2013.

[25] R. O. Hassan, M. Abdelhalim, and S.-D. Habib. Reliable pre-scheduling delay estimation for hardware/software partitioning. pages 1246 – 1250, Columbus, OH, United states, 2013.

[26] A. Srivastava, S. Vijay, A. Negi, P. Shrivastava, and A. Singh. DTMF based intelligent farming robotic vehicle: An ease to farmers. pages 206 – 210, Coimbatore, India, 2014.

[27] P. Asthana and S. Mangesh. Capacitor less dram cell design for high performance embedded system. Pages 554 – 559, Delhi, India, 2014.

[28] L. Yu, T. Chen, M. Wu, and L. Liu. Buffer on last level cache for cpu and gpgpu data sharing. pages 417 – 420, Paris, France, 2014.

[29] K. Prashanth, P. S. Akram, and T. A. Reddy. Real-time issues in embedded system design. pages 167 – 171, Andhra Pradesh, India, 2015.

[30] Z. Cheng, H. Zhang, Y. Tan, and A. O. Lim. Dpsc: A novel scheduling strategy for overloaded real-time systems. pages 1017 – 1023, Chengdu, China, 2015.

[31] A. Ghorbel, N. Ben Amor, and M. Jallouli. An embedded real-time hands free control of an electrical wheelchair. pages 221 – 224, Valletta, Malta, 2015.

[32] M. Kovatsch, S. Duquennoy, A. Dunkels, "A Low-Power CoAP for Contiki," Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on , vol., no., pp.855,860, 17-22 Oct. 2011 doi: 10.1109/MASS.2011.100

[33] Vladimir Dyo. "Evolution and sustainability of a wildlife monitoring sensor network." Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems. ACM, 2010.

[34] B. Zhu, A. Joseph, S. Sastry. "A Taxonomy of Cyber Attacks on SCADA Systems," Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing , vol., no., pp.380,388, 19-22 Oct. 2011 doi: 10.1109/iThings/CPSCom.2011.34

[35] Antonio Luis Carodoso Silva, Raimundo Santos Moura and Naziozênio Antônio Lacerda. “Structure Analysis of the Abstracts of Scientific Papers via Graphs”. 2013. 9 pages. Academic Report - Federal University of Piauí, Teresina, Brazil.