Automated Framework for Evaluating Sustainability and Resilience in Higher Education Curriculum

Madhuri Kumari, Mamta Mehra, and Dieter Pfoser

1 Introduction

With the advances in technology, sustainable development is gaining great attention. The 2030 agenda for sustainable development [11] motivates us to take transformative steps for shifting the world towards a sustainable and resilient path. The operationalization of sustainability and resilience requires specific skill sets. Looking at the demand of the time, it is imminent that the higher educational institutes (HEIs) need to equip the graduates with required skills.

Sustainability refers to the development of products and services which do not limit the future generations need or ability to acquire the same product and services [2]. The concept can be related to societal, economic, cultural, technological or psychological practices thus qualifying sustainability as social sustainability, economic sustainability, cultural sustainability, technological sustainability and psychological sustainability [5] respectively. Resilience is used to identify products that do not get damaged and deformed easily and can be used for a long time. Resilience is the ability of any system to prepare for facing threats, to absorb impact, to bounce back and adapt to challenges, adversity and stressful situations [7].

M. Kumari (✉)
Department of Civil Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India
e-mail: mkumari@amity.edu

M. Mehra
District of Columbia, Global Council for Science and the Environment, Washington, Columbia, USA

D. Pfoser
George Mason University, Virginia, USA

D. Bajaj et al. (eds.), *Sustainable Built Environment*, Lecture Notes in Civil Engineering 451, https://doi.org/10.1007/978-981-99-8842-6_1

As part of quality education, the fourth sustainable goal of the UN-SDGs, sustainability and resilience related topics should be covered in the university course offerings. The curriculums are the foundation of academic learning and practical experiences in higher education. It is required to include sustainability and resiliency related courses as part of curriculum for a future ready academic framework. In order to develop and design new courses, the reconnaissance of existing curriculum system has to be carried out to understand the current state of course offerings in context of sustainability and resiliency.

With increasing number of universities, institutions and courses across the globe, manual review of the existing system is not feasible. Further, no technology support is available in market for automated review of the course catalogues. Thus, the curriculum review process is manual and very tedious. This provides an opportunity to develop a tool and automate the curriculum review process. Considering the advances in web crawling and data analytics technology, this study proposes an automated framework for evaluating the curriculums of HEIs. This tool can read the contents of course catalogue from the website of universities and further perform data collection and analysis of the courses. As a final outcome, this tool lists the courses in the area of sustainability and resilience and provides insight into the type of the courses. The data collected from the automated system can be further analysed using computational method of text analytics and keyword ontology. The analysis result available from this tool can be used to identify the gap areas and then develop appropriate courses in the area of sustainability and resilience.

2 Literature Review and Research Gap

Many researchers have supported the need for automation of several aspects of education system for getting right input and feedback required for timely improvisation of the system. In a higher educational institute, using of digital technology for automating various educational processes will not only help in achieving a high quality of education but will also help decision makers in monitoring and expediting the changes in HEI as per market demand. At global level, summative assessments are carried out to provide feedback and data on general trends and the state of the education system [10]. Digitalization is much discussed intervention in the education sector. In a study presented by Wright et al. the author has discussed the use of real-time data for adaptive management in education [12]. This study also presents the advantage of such real-time data management system for data-informed decisions helping in teaching and learning experiences. Several works have been carried out in automating the process of the educational institutions. For the past decade, web scraping and web crawling have witnessed application in tasks involving mining the data from internet sources [1, 6, 13]. According to Suhirman et al. [9], the use of data mining or web scraping in education can no longer be considered a new methodology. It is known to be a powerful tool that can render significant results in curriculum design. However, it is a significantly underused methodology in curriculum design

research [8]. Automation framework for accreditation of HEI [3], converting the collected data into valuable knowledge, student performance assessment and web-based platform for educational processes are some of the preferred areas by the researchers.

These studies collectively highlight the growing interest in automating the assessment of resilience and sustainability courses offered by universities. They emphasize the benefits of using text analysis, machine learning, and data-driven approaches to streamline the evaluation process and ensure consistent standards for these crucial aspects of higher education.

To assess the performance and benchmark HEIs, ranking and accreditation has an important role to play [5]. Globally there are many recognized ranking and accreditation, and curriculum review forms an integral module of these process. However, there are limited studies or research work on automation of curriculum review due to the inconsistent format of the course catalogue of universities.

Considering the outlined research gap in the above paragraph, there is a scope for developing automation tool that can check and review the course catalogue and provide insights and future course of actions. Through this study, authors have proposed an automated framework for evaluating the curriculums of HEIs for sustainability and resilience. This framework is modular and can be expanded for any future requirements.

3 Study Data Set

This study was supported by Global Council for Science and the Environment (GCSE), USA. The organization of GCSE is a credible source of data and analysis of trends in higher education landscape, especially related to sustainability education and environmental studies. The dataset used in the study was divided into a training dataset and validation dataset. The list of 165 member universities of GCSE as shown in Fig. 1 was considered as training dataset which was used to train and develop the proposed automation framework. A comprehensive list of 2000 universities of USA was used to validate the developed framework.

4 Methodology

The automation framework was developed using four step approach as outlined in Fig. 2. As a first step, the training dataset was thoroughly analysed by visiting the web pages of each of these 165 GCSE member universities. The website of the universities was traversed to find the catalogue page. Thus, the catalogue URLs were listed corresponding to the list of universities. In the second step, the catalogue page was explored to identify the pattern of course information arrangement. This step was critical for framing the strategies and algorithms for scraping the course related

Fig. 1 Training dataset (: Location of Universities)

data from catalogue web page. In the third step, the normalization of web response from catalogue web pages of different universities was done to list down the set of common information available across most of the catalogue information. This information formed the base for the schema designed for the extraction of course related data. As a final step, automated framework was developed using web crawling technologies.

1. **Identify catalogue URLs**: In this stage, the websites of 165 GCSE member universities were traversed to check the placement of course catalogue information and extraction of course catalogue URLs. were explored to understand the placement of course catalogue information. Subsequently, the URLs for accessing the course catalogue were identified. Figure 3 shows the representative data of course catalogue URLs collected in the first stage.

2. **Analyse catalogue page for identifying the patterns**: The presentation and arrangement of course catalogue information on web pages varies across different universities. In general, the course catalogue contains a list of courses which are

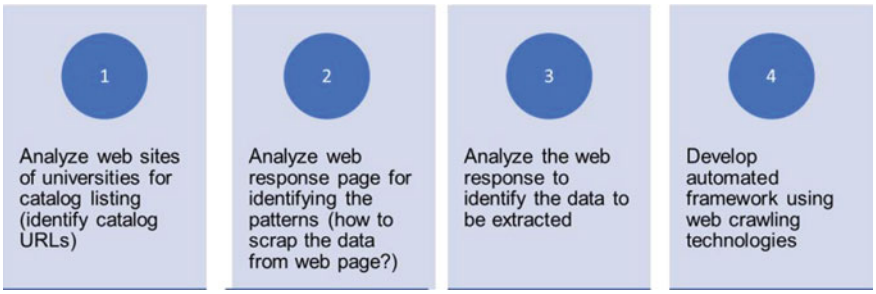


Fig. 2 Methodology

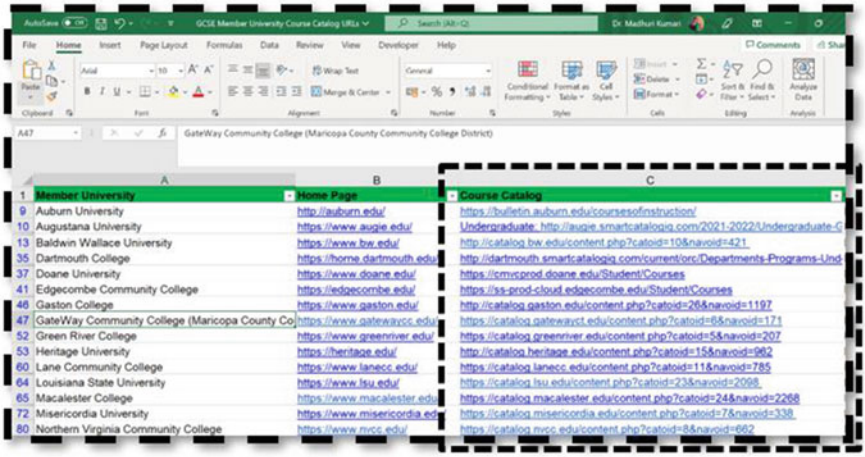


Fig. 3 Identification of course catalogue URLs

either arranged semester wise, UG/PG basis, department wise or presented as a single consolidated list with pagination. To develop web crawler, it is required to identify patterns of information arrangement on webpage. In this stage, an agile approach was used for identification of pattern and sub pattern which evolved as development progressed. In stage 1, the pattern at macro level was identified as table, non-table and pdf as shown in Fig. 4.

In stage 2, the high-level pattern was further nailed down and patterns for accessing details of courses was identified as (a) major divided directory: The courses were listed alphabetically and traversed using menu 'Courses A-Z'. The course name was displayed as hyperlink which on click opens a new page with course details. (b) full listing: All the courses offered by the university are listed on single page with course details presented in tabular format. (c) search-based directory: The subset of courses offered by the university is shown on the web page based on the search criteria. Figure 5 shows stage 2 of pattern identification.

In stage 3, the most common pattern of major divided directory and full course listing was further explored and a micro-level pattern identification was carried out which can form as an input for designing web crawler and parser. The micro-level pattern was identified as (a) major divided directory with details of course opening in new page (b) major divided directory with details of course opening on the same page (c) full course listing with all course information on single page displayed at single click of course catalogue menu. Figure 6 shows stage 3 of pattern identification at micro-level.

3. **Analyse the web response to identify the data to be extracted**: The course information common across different web response pages was identified. It was found that most web pages render the course catalogue with each course qualified by course title (along with course code), course description and course url that

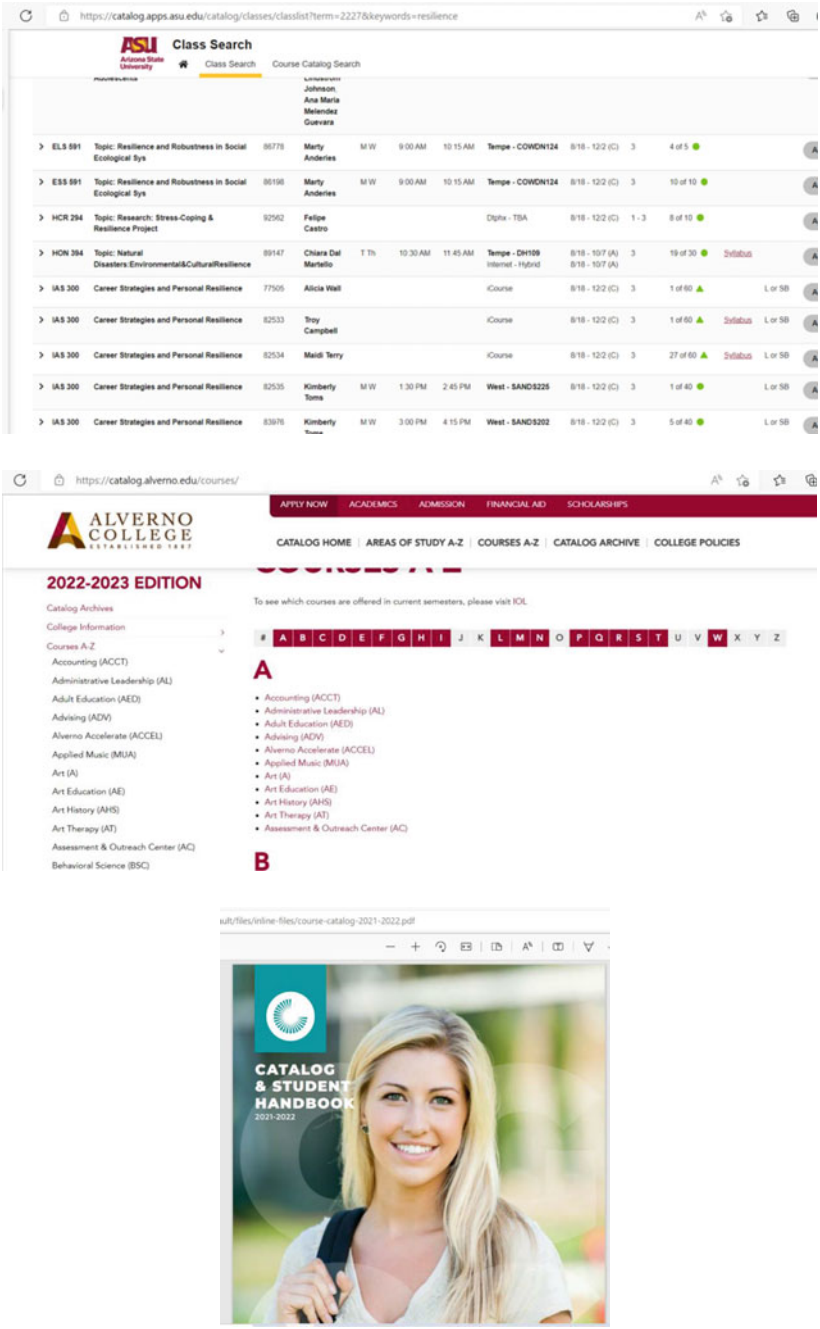


Fig. 4 Stage 1 of pattern identification—macro—**a** Table **b** Non-Table, **c** PDF [14–16]

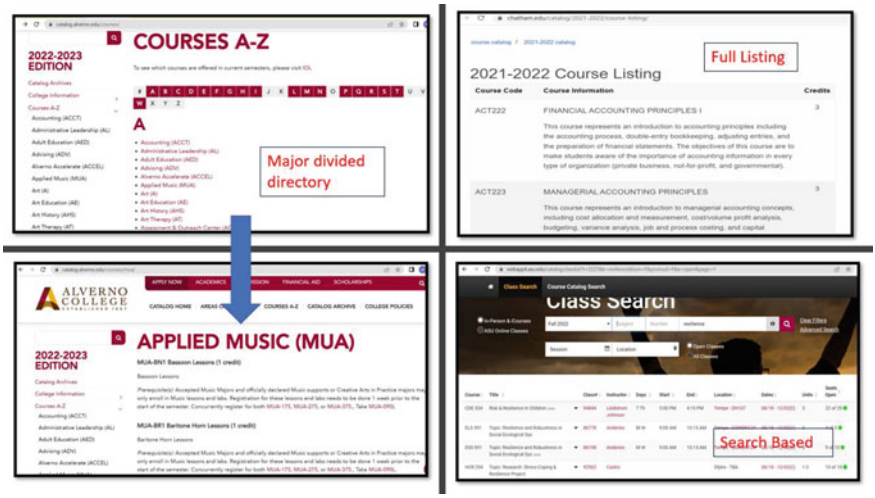


Fig. 5 Stage 2 of pattern identification (macro level—major divided directory, full listing, search-based) [14–16]

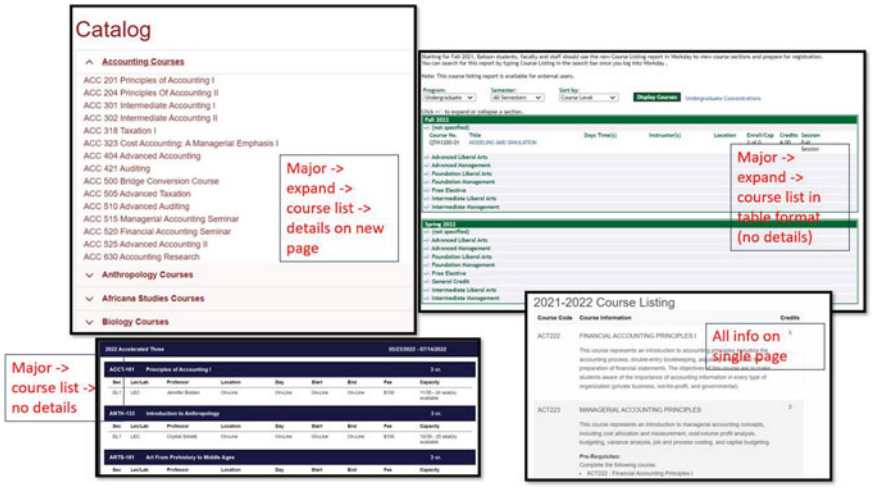


Fig. 6 Stage 3 of pattern identification (micro level—major divided directory with details on new page, major divided directory with inline course details, full course listing) [14–16]

points to course details. Based on the finding the parser was designed to extract course code and course title, course description and course urls (Fig. 7).

4. **Develop automated framework**: An automated web crawler-based framework was developed using open-source technology which can read the contents of course catalogue from the website of universities and further parse the webpage,

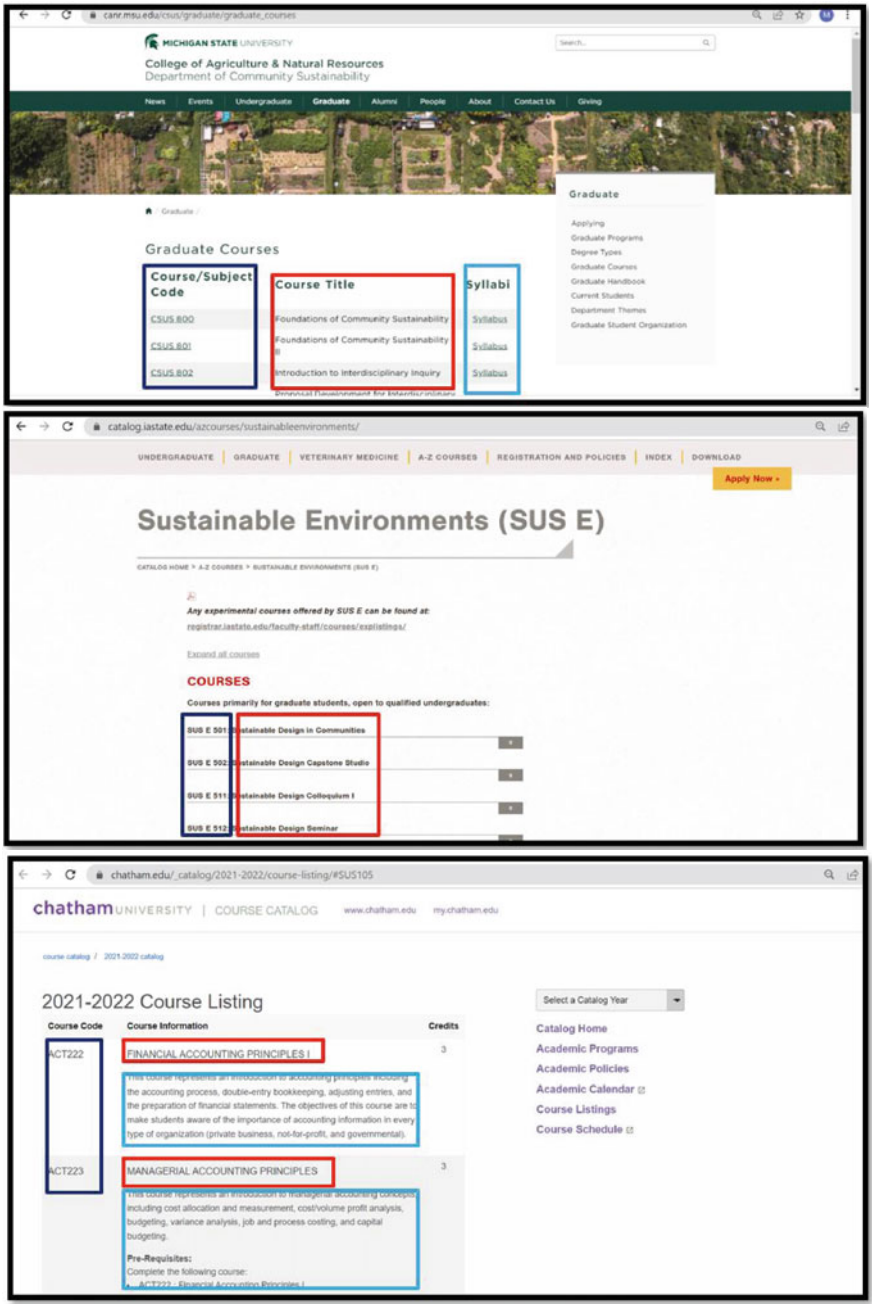


Fig. 7 Common course information across different catalogue pages [14–16]

perform data collection and analysis of the courses. The system was developed for a subset of universities in United States. It was tested using the validation dataset containing the comprehensive list of universities in United States.

5 Automated Framework for Evaluating Sustainability and Resilience in Higher Education Curriculum

The automated framework consists of four important components namely, web driver, web crawler, web response parser and output formatter as shown in Fig. 8. The framework was developed using python. The web driver was responsible for reading the URL of course catalogue and sending the request to the URL. It used selenium library of python. The web crawler was used for collecting the response from the website and storing it offline as a text file. This approach of dumping response in file without parsing was helpful in saving time required for revisit of the website and handle bulky response from the website. The web parser was developed using beautifulsoup library of python. This component was responsible for parsing the web page and extracting the course related information as per designed response schema and storing it in object. The web parser passed on the control to output formatter synchronously. The output formatter prepared JSON object and dumped it in output text file which was then stored offline. The technology stack used for the development of this automated framework is shown in Fig. 9. Google colab was used as the platform to write and execute the python code.

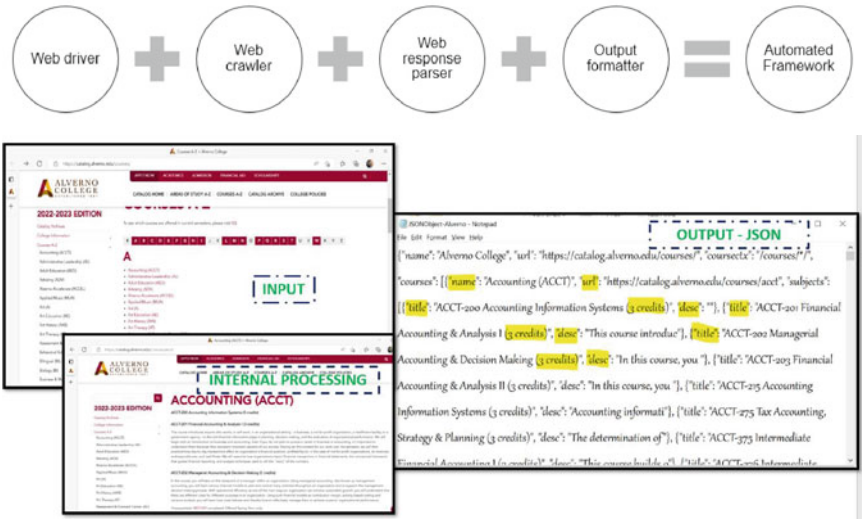


Fig. 8 Automated framework for evaluating sustainability and resilience in higher education curriculum []

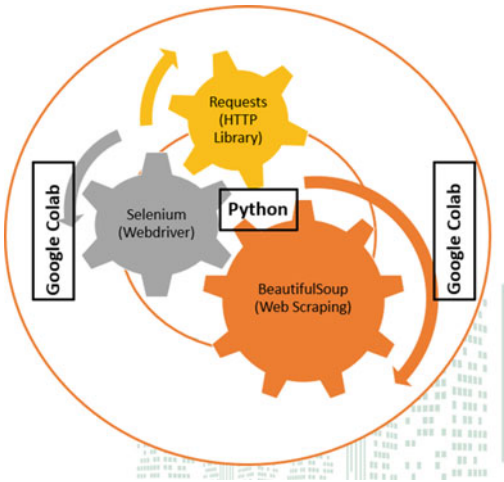


Fig. 9 Technology stack used for developing Automated framework

In summary, this framework takes catalogue URL of a university as input and generates a well formatted file with course information extracted from the course catalogue web page. The output is in popular data-interchange format, JSON which is lightweight, text-based and represents the data as a collection of name/value pairs. It can be easily consumed by any programming language and supports ease of integration.

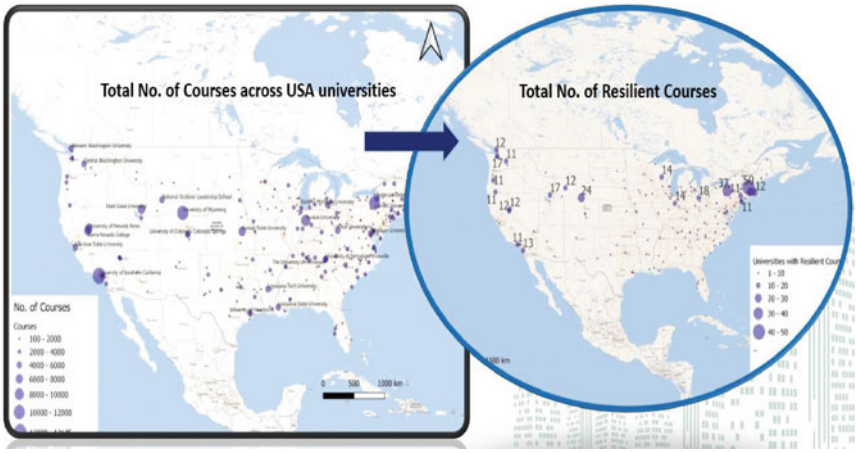
6 Result and Conclusion

The course data collected through the web scraping programme was analysed to understand the integration of sustainability and resilience in the courses being offered across the universities. The data so collected from the automated system was analysed using computational method of text analytics and keyword ontology. Text mining techniques were applied for distilling insights from the collected course data.

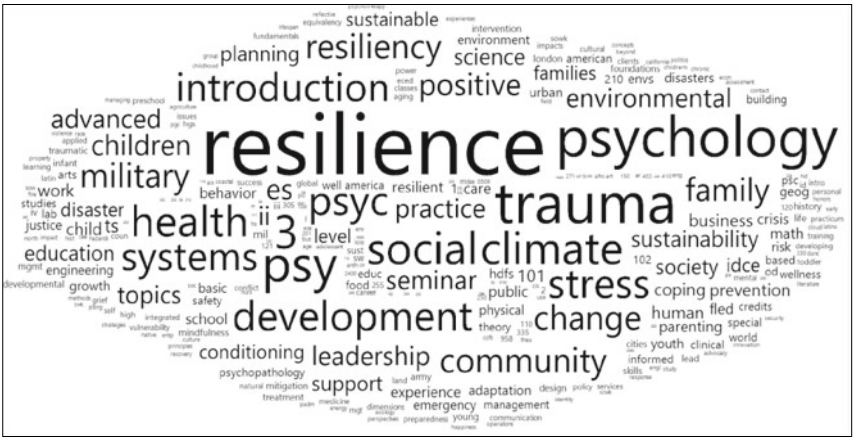
The field of course title and course description were searched for keywords that directly or indirectly indicated the relevance with regard to resilience. The keywords used as a direct indicator were ‘resilient’, ‘resilience’, ‘resiliency’, ‘sustainable’, ‘sustainability’. The keywords used as an indirect indicator were ‘energy’ and ‘climate’. This was followed by analysing the search result further for finding the resilient courses in the area of environment, energy and climate. Word frequency analysis was carried out on the course title and course description field of search results to understand the occurrence of keywords in resilience related courses. The frequency of the words was used to infer the top 5 areas of resilient ready courses. To make semantic analysis more accurate in text mining, collocation analysis using the Bag of Words (BoW) approach was applied. The course title and course description were tokenized using 2-g, 3-g and 4-g to find the occurrence of environment, energy or climate in the courses matching the resilience criteria.

As a final outcome, this tool listed the courses in the area of sustainability and resilience based on the keywords used for searching. The data can be used for getting an insight into the type of the courses as presented in Fig. 10. The top 10 phrases using 2-g, 3-g and 4-g.

The spatial map of resilience courses as in Fig. 10a clearly shows that less number of resilient or sustainability related courses are being offered by the universities across USA. East Coast of USA has better coverage of resilience in the courses as compared to the West Coast of USA. The word cloud of course title having direct mention of resilience or sustainability in the title is plotted in Fig. 10b. It shows that there are varied areas covered as part of resilience education. The top 10 areas include



(a) Spatial representation of total courses versus resilience courses on the map



(b) Word cloud representing the frequency of occurence of different words in the resilience course title

Fig. 10 Spatial map and word cloud for resilience courses

psychology, social, climate, health, community, development, environmental, trauma and stress. The word cloud can be supplemented by insights from Table 1 which provide top phrases using 2-g, 3-g and 4-g, indicates that community resilience is most popular course across the university. Climate, weather, food security are also popular resilience subjects. Psychology courses covering trauma and stress resilience are also offered in many of the universities of USA. Resilience in security, risk, system reliability is also covered as course offering in few universities. Though the universities across USA are offering courses related to resilience, still the knowhow of the students in this matter is very limited. This clearly indicates the need for designing and introducing new courses in areas of resilience and sustainability. The courses related to resilience in the field of environment, disaster need to be designed and should be offered in all the universities so that students are aligned to the SDGs.

This outcome can be effectively used to identify the gap and thus preparing the list of courses to be developed and designed for an efficient and future ready curriculum.

Table 1 Top phrases using Bag of Words approach and tokenized using 2-g, 3-g and 4-g

2 words phrase	4 words phrase
Community resilience	Agroecology and resilient communities
Resilient communities	climate change community resilience
Building resilience	Family stress and resilience
Coastal resilience	Stress and resiliency
Community resiliency	Building community environmental resilience
Environmental resilience	Climate and weather resiliency
Resilient cities	Food security and resilient
Resilient systems	Security and resilient communities
Weather resiliency	System reliability and resilience
3 words phrase	
Trauma and resilience	Legacy of resilience
Agroecology and resilient	Reintegration and resilience
Change community resilience	Reliability and resilience
Stress and resilience	Resilience and freedom
Identity and resiliency	Resilience and recovery
Resilience in children	Resiliency and coping
Resiliency in public	Resiliency in meteorology
Risk and resiliency	Resilient communities practicum
Stress and resiliency	Risk to resiliency
Sustainability and resiliency	Security and resilient
Threats and resilience	Skills of resilience
Community environmental resilience	Trauma and resiliency

7 Challenges and Future Scope of Work

The current course offering in various universities is varied, with each college offering different courses under sustainability and resilience. These varied courses need some normalization so that they can be compared and evaluated. It is also important to grade the various courses so that no redundant courses are being offered.

The dynamically changing pages of catalogue presented a challenge during framework development. With new semester and academic session, pointer to catalogue changes. The catalogue URL working today may not be working in future. Also, the web response of the given catalogue URL may change in future. Considering the cases analysed during the development, around 50–65% cases of web response were not generalizable and hence developing a fool proof framework is difficult. In many cases, the university maintains the course catalogue in the form of a pdf copy, parsing of which is not handled as part of our study and forms a limitation for usage of the framework.

Several technological challenges were encountered during development of the framework. Initially, the four components of the framework worked synchronously, and the control serially passed on from one component to another. However, the http request timed out while fetching multiple pages of web response as many web pages of catalogue presented the course list along with pagination. Further, the large size of web response could not be stored in memory and hence the data processing was interrupted. To handle this timeout error, the http request was made asynchronous with increased time lag and the web response was cached in a file. The cached response was advantageous as revisit of web page could be avoided thus saving time.

The developed framework can be extended and generalized in future to cater to higher education institution across the world. Further, the framework is flexible to handle the queries related to finding courses in any domain provided the keyword is defined. This framework can be used for identification of university readiness for subjects beyond sustainability and resilience. Also, there is a scope of increasing the data collection from catalogue for better report generation and handling complex queries. The framework can be considered as a base for incorporating more customization to increase the reach of this framework to recruiters, research institutions and policy makers.

Acknowledgements The authors would like to acknowledge the funding support provided by the Sloan Foundation, Department of Energy, United States, and George Mason University. The corresponding author would also like to acknowledge GCSE for providing the opportunity to conduct the study and develop the automation framework.

References

1. Adelola MA, Adewale OS, Iwasokun GB (2016) Graph theoretic and genetic algorithm-basedmodel for web content mining. Int J Comput Sci 13(6):68–75.https://doi.org/10.20943/01201606.6875
2. Aminpour P, Gray S, Richardson R, Singer A, Castro-Diaz L, Schaefer M, Chikowore NR (2019) Perspectives of scholars on the nature of sustainability: a survey study. Int J Sustain Higher Education
3. Alhiyafi J, Abuaqel I (2018) An Automated System for Collection and Analysis of ABET Curriculum Requirements. International Conference on Computational Science and Computational Intelligence (CSCI), pp 663–666. https://doi.org/10.1109/CSCI46756.2018.00133
4. Di Fabio A, Kenny ME (2018) Connectedness to Nature, Personality Traits and Empathy from a Sustainability Perspective. Curr Psychol 1–12. https://doi.org/10.1007/s12144-018-0031-4
5. Fernandes JO, Singh B (2022) Accreditation and ranking of higher education institutions (HEIs): review, observations and recommendations for the Indian higher education system. The TQM Journal 34(5):1013–1038. https://doi.org/10.1108/TQM-04-2021-0115
6. Kimmons R (2017) Open to all? Nationwide evaluation of higher-priority web accessibility-considerations among higher education websites. Journal of Computing in HigherEducation 29(3):434–450. https://doi.org/10.1007/s12528-017-9151-3
7. Marchese D, Reynolds E, Matthew B, Morgan H, Clark SS, Linkov I (2018) Resilience and sustainability: Similarities and differences in environmental management applications. Science of The Total Environment, Vol 613–614, 1275–1283, ISSN 0048–9697. https://doi.org/10.1016/j.scitotenv.2017.09.086.
8. Moretti A, González-Brenes JP, McKnight K (2014) Data-driven curriculum design: Mining theweb to make better teaching decisions. Proceedings of the 7th International Conference onEducational Data Mining (EDM 2014). Retrieved from https://files.eric.ed.gov/fulltext/ED558339.pdf
9. Suhirman S, Herawan T, Chiroma H, Zain JM (2014) Data mining for educationdecision support: A review. Int J Emerg Technol Learn 9(6):4–19
10. World Bank, UNICEF & UNESCO (2020). The state of the global education crisis: a path to recovery. https://www.unicef.org/media/111621/file/%20The%20State%20of%20the%20Global%20Education%20Crisis.pdf%20.pdf
11. https://sdgs.un.org/2030agenda
12. Wright E, Osborne S, Aggarwal M (2022) Digital tools for real-time data collection in education, The Brookings Institution. https://www.brookings.edu/essay/digital-tools-for-real-time-data-collection-in-education/
13. Raicu I (2019) Financial banking dataset for supervised machine learning classification. Informatica Economica, 23(1), 37–49. https://doi.org/10.12948/issn14531305/23.1.2019.04
14. https://www.canr.msu.edu/
15. https://catalog.iastate.edu/azcourses/sustainableenvironments/
16. https://chatham.edu/_catalog/2021-2022/course-listing/#SUS105